



UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
FACULDADE DE MEDICINA
DEPARTAMENTO DE ANATOMÍA PATOLÓXICA E CIENCIAS FORENSES

Identificación de variantes de riesgo en regiones reguladoras de genes asociados a esquizofrenia

Memoria para optar al grado de Doctor

presentada por:

Javier González Peñas

Marzo 2016





El Doctor Ángel Carracedo Álvarez, Catedrático de Medicina Legal de la Facultad de Medicina de la Universidad de Santiago de Compostela, y el Doctor Javier Costas Costas, Investigador del Servicio Gallego de Salud en el Instituto de Investigación Sanitaria de Santiago de Compostela.

CERTIFICAN:

Que el presente trabajo que lleva por título “Identificación de variantes de riesgo en regiones reguladoras de genes asociados a esquizofrenia”, realizado por Javier González Peñas bajo nuestra dirección, ha sido revisado y está en disposición de ser presentado para optar al grado de Doctor en Medicina molecular.

Y para que así conste, firmamos el presente certificado en Santiago de Compostela, a 23 de Marzo de 2016

Fdo: Dr. Ángel Carracedo Álvarez

Fdo: Dr. Javier Costas Costas



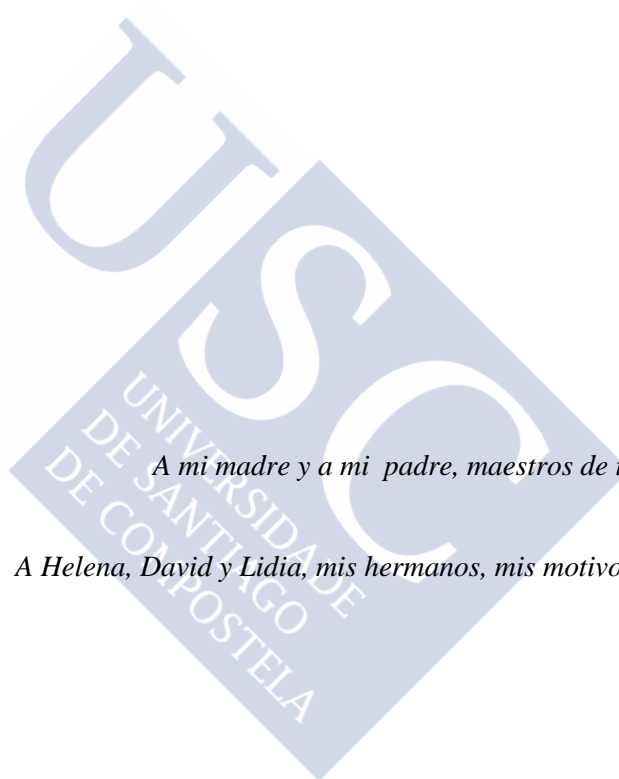
Resumen

La esquizofrenia es una enfermedad de elevada prevalencia en la población (0,5%), altamente incapacitante y de una naturaleza multicausal, lo que provoca que su diagnóstico sea complejo. Los estudios de heredabilidad más representativos han situado la varianza genética asociada a la enfermedad en unos valores de 65-80% del total, por lo que se trata de uno de los trastornos complejos más genéticos que existen. No obstante, los múltiples estudios realizados para describir la naturaleza genética del mismo han conseguido explicar únicamente un porcentaje muy reducido de la heredabilidad asociada cuando se han fundamentado en variación común, mediante estudios de asociación de genoma completo (GWAS), o variación rara, asociada a regiones exónicas mediante estudios de secuenciación de miles de exomas. En la dicotomía existente entre las hipótesis genéticas que estudian la variabilidad asociada a la esquizofrenia basándose en variantes raras o comunes, todavía existe una gran ausencia de trabajos que analicen las primeras en el ámbito de las regiones reguladoras de la expresión génica, y mayoritariamente no codificantes. De hecho, los pocos trabajos al respecto sugieren un papel muy relevante de estas regiones en la etiología de la enfermedad.

El presente proyecto de tesis se ha centrado en el estudio de la variación genética en regiones reguladoras, seleccionadas en base a patrones extremos de conservación filogenética o apoyándose en datos generados por proyectos masivos centrados en el estudio de la funcionalidad global del genoma como el proyecto Encode o el proyecto Roadmap, así como en otros enfocados en la distribución de la expresión génica a lo largo de las distintas etapas del neurodesarrollo y en las diferentes áreas del cerebro, como el proyecto Brainspan. De este modo, en primer lugar, se estudió la variación genética rara en el elemento ultraconservado no codificante y *enhancer* de la expresión génica UC435 del gen *TCF4*, un factor de transcripción de la familia bHLH ampliamente asociado en los estudios más relevantes, revelándose una ausencia de variación reguladora en esquizofrenia a lo largo del mismo. En segundo lugar, se realizó un análisis de asociación de la variación genética presente en población europea en la base de datos de los 1000 genomas a lo largo de las 14 regiones conservadas en mamíferos pero muy aceleradas en humanos, y con funcionalidad de *enhancer*, presentes en *NPAS3*, el gen con mayor número de las mismas y además, asociado a esquizofrenia y a trastornos del neurodesarrollo en numerosos trabajos previos. Al igual que en *TCF4*, el estudio de la variación genética en regiones reguladoras de *NPAS3*, otro factor de transcripción de la familia de bHLH, no demostró la existencia de asociación al respecto, sugiriendo que la estrategia de análisis de variación reguladora basada en patrones de conservación genética no es la más adecuada en el estudio de la esquizofrenia. En tercer lugar, se analizó de forma sistemática, mediante resecuenciación dirigida de nueva generación, la variación presente en las regiones promotoras

de genes asociados a esquizofrenia a raíz de estudios de variaciones en el número de copias (CNV) o de GWAS, demostrándose un enriquecimiento significativo de variación reguladora en estas regiones en casos. Además, la subdivisión de las mismas en estados de la cromatina basados en modificación epigenéticas, reveló la presencia de 9 variantes en casos y ninguna en controles a lo largo de los promotores de los genes *HNF1B* y *LHX1*, pertenecientes a la región 17q12, y con una funcionalidad de represor polycomb. El análisis en profundidad de la evolución de estos estados funcionales a lo largo del neurodesarrollo sugiere que estas regiones se hallan inactivas a partir de estadios muy iniciales de diferenciación celular, por lo que dichas variantes podrían modificar la estructura de la cromatina en etapas muy primordiales del desarrollo y la alteración en la expresión de los genes que regulan en éstas ser un elemento de alto riesgo para la aparición, años más tarde, de la esquizofrenia, en consonancia con las numerosas evidencias que sugieren la hipótesis del neurodesarrollo como explicación del origen de la enfermedad. En último lugar, mediante secuenciación de ARN de cerebro *post mortem*, se estudiaron las diferencias de expresión y la variación genética presentes entre grupos de casos y controles bien caracterizados. Dicho análisis demostró, mediante la utilización de módulos de coexpresión y perfiles de expresión génica en líneas celulares cerebrales descritas en estudios previos, la existencia de una subexpresión génica de aquellos genes específicos de líneas celulares neuronales, así como la sobreexpresión de aquellos más específicos de astrocitos, independiente de la medicación antipsicótica administrado a lo largo de la vida de los pacientes. El estudio de la variación genética rara relativa a los módulos de coexpresión diferencialmente expresados reveló una acumulación en casos de variantes en genes pertenecientes a módulos coexpresados en esquizofrenia y enriquecidos en genes de astrocitos. Además, también se demostró que las diferencias de expresión existentes en estos genes pueden ser explicadas, al menos en parte, por la acumulación de estas variantes, reforzando el papel de los astrocitos en la etiología de la enfermedad y demostrando que su implicación va más allá de un efecto secundario, pudiendo ser más una causa que una consecuencia de la enfermedad.

Los resultados expuestos a lo largo de los trabajos presentados en este proyecto de tesis reivindican una mayor consideración de las regiones reguladoras de la expresión génica para la búsqueda de factores genéticos de susceptibilidad a esquizofrenia y, de forma más relevante, aquellos que afectan a estados funcionales descritos en función de características epigenéticas y a genes específicos de líneas celulares determinadas o áreas del cerebro importantes en momentos críticos del neurodesarrollo.



A mi madre y a mi padre, maestros de todo lo que soy.

A Helena, David y Lidia, mis hermanos, mis motivos más sinceros.



Agradecimientos

En primer lugar, me gustaría agradecer a Ángel Carracedo la posibilidad que me dio, hace ya unos cuantos años, de formar parte de su grupo de investigación. En un momento en que quise dar un giro importante a mi formación, me mostró su apoyo y disponibilidad para ayudarme a llevarlo a cabo. Admiro su energía para impulsar la investigación en Galicia, a pesar de las enormes dificultades, y su incansable motivación.

Quiero agradecer a Xabi Costas, mi tutor durante estos años de doctorado, la intensa formación recibida, siempre desde la más admirable ética profesional y honradez, así como su labor al frente del grupo de psiquiatría genética. Creo y espero que el rigor científico con el que lleva a cabo cada uno de sus proyectos sirva de ejemplo para muchos, entre los que me incluyo, y ayude a hacer de esta nuestra profesión algo más digno y respetable, para que alcance el lugar que se merece.

Esta tesis no habría sido lo mismo sin la ayuda del *tridente catacroker*: Julio, Xaquín y Cris. Tengo que dar las gracias a los intensos debates en los que construimos un mundo de muchachada, hipsters, sociopolítica, música, y un largo etcétera que ha contribuido, sin ninguna duda, a que mi mente sea más científica. La ciencia nunca debería olvidarse de la creatividad.

Gracias también a Eva, Jorge, Joja, Juan,... y a todos los miembros de los grupos de Medicina Xenómica que me han ayudado y apoyado para que este día tuviese lugar.

Al otro lado del charco, mención especial para Bin Xu, mi tutor durante unos inolvidables 6 meses en Nueva York, y a todos los amigos y compañeros de aventuras de más allá del charco con los que pude disfrutar tanto personal como profesionalmente una enorme cantidad de experiencias. Gracias a la Fundación Barrié por hacerlo posible, en un momento donde nadie más me dio la oportunidad.

Quiero agradecer a mis padres y hermanos su apoyo incondicional. Nada de lo que soy habría sido posible sin su ejemplo diario, sin ellos. Gracias por enseñarme a mejorar y a valorar el esfuerzo personal, y más todavía, por inculcarme sus valores.

Gracias a Gabrielle, Braisótopo, al Luko, The Tonate, Pete the dick, Giorgos, al *loco Lo Faunè* y a un muy largo etcétera que incluye a todos mis amigos y compañeros de vivencias extralaborales, con los que tengo la suerte de compartir Mesa, Magia, Mar, Montaña y Música.

Finalmente, quiero agradecer a todos aquellos que colaboran aportando y recogiendo muestras para que nuestro trabajo sea posible. De igual modo, no quiero pasar por alto la oportunidad de agradecer profundamente la labor de héroes modernos como Aaron Swartz o

Alexandra Elbakyan, que luchan o lucharon por la libre difusión del conocimiento científico, anteponiendo la libertad de información al cumplimiento de leyes injustas.

Esta tesis fue realizada con la ayuda de un contrato predoctoral de la Xunta de Galicia, un convenio de la Fundación María José Jove y una beca de la Fundación Barrié; y estuvo financiada a través de proyectos de investigación por el Instituto de Salud Carlos III (CP11/00163 and PI11/00770). Así mismo, los resultados fueron posibles gracias a la colaboración con las plataformas tecnológicas del CEGEN-ISCIH, de la Fundación Pública Galega de Medicina Xenómica y del Centro de Supercomputación de Galicia (CESGA).





*“También estaba el cura.
Llegó a tiempo de confesarme que Dios no existía.”*



ÍNDICE

1. INTRODUCCIÓN	1
1.1. ASPECTOS CLÍNICOS Y NEUROBIOLÓGICOS DE LA ESQUIZOFRENIA	3
1.1.1. Antecedentes históricos	3
1.1.2. Esquizofrenia. Definición y sintomatología	6
1.1.3. Curso de la enfermedad	9
1.1.4. Epidemiología	12
1.1.5. Neuroanatomía y alteraciones estructurales	15
1.1.6. Neurobiología de la esquizofrenia	17
1.1.6.1. Hipótesis neuroquímicas	17
1.1.6.2. Hipótesis inflamatoria	19
1.1.6.3. Hipótesis del neurodesarrollo	22
1.1.6.4. Modelo etiológico mixto	24
1.2. GENÉTICA DE LA ESQUIZOFRENIA	25
1.2.1. Epidemiología genética. Heredabilidad de la esquizofrenia	26
1.2.2. Estudios genéticos en esquizofrenia	30
1.2.2.1. Estudios de ligamiento	31
1.2.2.2. Estudios de asociación en genes candidatos	35
1.2.2.3. Estudios de asociación de genoma completo (GWAS)	44
1.2.2.4. Detección de Variaciones en el número de copias (CNV)	50
1.2.2.5. Next Generation Sequencing (NGS) y estudios de secuenciación en esquizofrenia.	56
1.2.3. Arquitectura genética de la esquizofrenia	64
1.2.3.1. Hipótesis variante común – enfermedad común Vs hipótesis variante rara – enfermedad común	65
1.2.3.2. Heredabilidad perdida	67

1.3. REGULACIÓN DE LA EXPRESIÓN GÉNICA EN LA ESQUIZOFRENIA	69
1.3.1. Estrategias para abordar un estudio de regulación de la expresión génica en esquizofrenia	69
1.3.2. Estudios de expresión en esquizofrenia	72
1.3.2.1. Técnicas en el estudio de la expresión génica.	72
1.3.2.2. Genes diferencialmente expresados (DE) en esquizofrenia	76
1.3.3. La importancia de la conservación genética en la identificación de elementos reguladores de la expresión génica.	78
1.3.3.1. Conservación genética en mamíferos	78
1.3.3.2. Elementos conservados del genoma humano	80
1.3.3.3. Elementos conservados en mamíferos y acelerados en el linaje humano	82
1.3.4. Regulación de la transcripción	86
1.3.4.1. Elementos funcionales de la regulación de la expresión génica. Proyecto ENCODE	86
1.3.4.2. Estados epigenómicos específicos de tejido. Proyecto Roadmap	91
1.3.4.3. Estudio de la variación genética en elementos reguladores	93
2. JUSTIFICACIÓN, HIPÓTESIS Y OBJETIVOS	97
3. MATERIALES Y MÉTODOS	103
3.1. MUESTRAS ANALIZADAS	105
3.1.1. Información de las muestras empleadas	105
3.1.1.1. Muestras de ADN	105
3.1.1.2. Muestras de ARN	106
3.1.2. Preparación de las muestras para su utilización	108
3.2. SELECCIÓN DE REGIONES REGULADORAS	109
3.2.1. Elementos ultraconservados en genes asociados a esquizofrenia. <i>TCF4</i>	109

3.2.2.	Selección de variantes en elementos acelerados en humanos (HAEs) de genes asociados a esquizofrenia. <i>NPAS3</i>	109
3.2.3.	Selección de genes de interés para la secuenciación de promotores mediante NGS	110
3.3.	DETECCIÓN DE MUTACIONES MEDIANTE HRMA	112
3.3.1.	Introducción teórica	112
3.3.2.	Amplificación del elemento UC435 de <i>TCF4</i> para el análisis de mutaciones mediante HRMA	114
3.3.3.	Análisis de variantes genéticas en UC435 mediante HRMA	116
3.4.	TECNOLOGÍA DE GENOTIPADO DE SEQUENOM	117
3.4.1.	Introducción teórica	117
3.4.2.	Genotipado de variantes en elementos acelerados en humanos (HAEs) de <i>NPAS3</i>	118
3.4.3.	Confirmación mediante genotipado de variantes de secuenciación dirigida en regiones promotoras	118
3.5.	SECUENCIACIÓN DE SANGER	119
3.5.1.	Amplificación del elemento UC435 de <i>TCF4</i> para la secuenciación de Sanger de las muestras con posibles variantes.	119
3.5.2.	Secuenciación de las muestras con posibles variantes en UC435	120
3.5.3.	Secuenciación para la confirmación de las 30 regiones promotoras diseñadas	121
3.6.	SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS). PLATAFORMA DE SOLID	121
3.6.1.	Introducción teórica	121
3.6.2.	Secuenciación de genes de estudios GWAS y de CNV en esquizofrenia para el análisis de variantes reguladoras	123
3.6.2.1.	Diseño de las regiones diana de secuenciación (<i>targeted resequencing</i>)	123
3.6.2.2.	Amplificación de las regiones diana de secuenciación (<i>targeted resequencing</i>)	130
3.6.2.3.	Secuenciación de las regiones promotoras (<i>targeted resequencing</i>)	131

3.7.	SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS). PLATAFORMA DE ILLUMINA	132
3.7.1.	Introducción teórica	132
3.7.2.	Secuenciación de ARN de muestras de cerebro <i>post mortem</i> en esquizofrenia.	134
3.8.	ANÁLISIS BIOINFORMÁTICOS	135
3.8.1.	Secuenciación dirigida de ADN de regiones promotoras	135
3.8.1.1.	Mapeo frente al genoma de referencia.	135
3.8.1.2.	Detección y filtrado de variación genética rara en promotores.	135
3.8.1.3.	Anotación funcional de la variación genética detectada	137
3.8.2.	Secuenciación de transcriptoma de giro cingulado en esquizofrenia	138
3.8.2.1.	Mapeo frente al genoma de referencia.	138
3.8.2.2.	Estudio de la expresión diferencial en giro cingulado en esquizofrenia	138
3.8.2.3.	Estudio de la correlación entre expresión génica y medicación administrada a lo largo de la vida en los individuos esquizofrénicos.	141
3.8.2.4.	Detección y filtrado de variación genética en transcriptoma de cerebro en esquizofrenia.	143
3.8.2.5.	Detección de variación genética en genes expresados consistentemente en giro cingulado	143
3.8.2.6.	Anotación funcional de variantes en transcriptoma de giro cingulado	145
3.9.	ANÁLISIS ESTADÍSTICOS	145
3.9.1.	Estudio de la variación genética en UC435 de <i>TCF4</i>	145
3.9.2.	Estudio de la variación genética en HAEs de <i>NPAS3</i>	145
3.9.3.	Estudio de la variación genética en regiones promotoras	147
3.9.4.	Estudio del transcriptoma de giro cingulado en esquizofrenia	147
3.9.4.1.	Análisis de datos de expresión mediante GSEA y CSEA	147
3.9.4.2.	Análisis de correlación entre medicación y expresión génica	148
3.9.4.3.	Variación genética en módulos de coexpresión	148

3.9.4.4. Estudio de la influencia de la variación genética rara de genes altamente expresados en los niveles de expresión génica	149
--	-----

4. RESULTADOS 151

ESTUDIO I. VARIACIÓN GENÉTICA RARA EN ELEMENTOS CONSERVADOS DE *TCF4* DE POTENCIAL REGULADOR 153

I. 1. Introducción	153
I. 2. Estudio de la secuencia de UC435 en mamíferos y de la variación genética descrita en población humana.	154
I. 3. Análisis de la variación genética en UC435 en esquizofrenia mediante <i>High Resolution Melting Analysis</i> (HRMA).	155
I. 4. Secuenciación de las muestras con variantes “dudosas” en UC435	157

ESTUDIO II. VARIACIÓN GENÉTICA EN ELEMENTOS ACELERADOS EN HUMANOS EN *NPAS3*. 159

II. 1. Introducción	159
II.2. Genotipado de variantes en <i>NPAS3</i>	160
II. 3. Análisis de variantes en <i>NPAS3</i>	161

ESTUDIO III. RESECUENCIACIÓN DE REGIONES PROMOTORAS DE GENES DE RIESGO EN ESQUIZOFRENIA. 165

III. 1. Introducción	165
III. 2. Cobertura de las regiones analizadas	166
III. 3. Detección de variantes	168
III. 4. Confirmación de variantes secuenciadas mediante genotipado en Sequenom	169
III.5. Comparaciones caso-control basadas en RegulomeDB	170
III.6. Comparaciones caso-control basadas en los estados de la cromatina (Proyecto Roadmap)	171

ESTUDIO IV. ANÁLISIS DE TRANSCRIPTOMA DE GIRO CINGULADO EN ESQUIZOFRENIA. ESTUDIO DE VARIACIÓN GENÉTICA Y DIFERENCIAS DE EXPRESIÓN EN CEREBRO 177

IV. 1. Introducción	177
---------------------	-----

IV. 2. Estudio de expresión en giro cingulado en esquizofrenia	178
IV.2.1. Análisis de genes con expresión detectable	179
IV.2.2. Estudio de correlación con medicación antipsicótica	180
IV.2.3. Estudio de enriquecimiento de los datos de expresión en grupos de genes mediante GSEA.	181
IV.2.3.1. Análisis de enriquecimiento en módulos de coexpresión en esquizofrenia	181
IV.2.3.2. Análisis de enriquecimiento en genes específicos de tipo celular	183
IV.2.3.2. Análisis de enriquecimiento en genes relacionados con la sinapsis	185
IV. 3. Genes diferentemente expresados (DE) en giro cingulado en esquizofrenia.	186
V.3.1. Descripción de los genes DE en esquizofrenia	186
IV.3.2. Expresión relativa de genes DE subexpresados y sobreexpresados en esquizofrenia	189
IV.3.3. Enriquecimiento de genes DE en tipos celulares y tejidos específicos durante el desarrollo	189
IV.3.3.1. Análisis de enriquecimiento en tipos celulares cerebrales	191
IV.3.3.2. Análisis de enriquecimiento en áreas cerebrales durante el desarrollo.	194
IV. 4. Análisis de variantes de genes altamente expresados en las muestras de ARN de giro cingulado.	197
IV. 4. 1. Análisis descriptivo de las variantes detectadas y de los genes mutados. Los genes mutados se encuentran subexpresados en esquizofrenia.	197
IV. 4. 2. Estudio de la variación genética en módulos subexpresados en esquizofrenia. Disrupción de genes implicados en el desarrollo cortical en casos de esquizofrenia.	200
IV. 4. 3. Estudio de la variación genética en módulos sobreexpresados en esquizofrenia. Acumulación de variantes raras en casos respecto a controles.	204
IV. 5. Estudio de la influencia de las variantes genéticas en la expresión. La variación en genes sobreexpresados explica diferencias de expresión entre casos y controles	207

5. DISCUSIÓN

5.1.	ESTUDIO I. VARIACIÓN GENÉTICA RARA EN ELEMENTOS CONSERVADOS DE <i>TCF4</i> DE POTENCIAL REGULADOR	213
5.2.	ESTUDIO II. VARIACIÓN GENÉTICA EN ELEMENTOS ACELERADOS EN HUMANOS EN <i>NPAS3</i>.	220
5.3.	ESTUDIO III. RESECUENCIACIÓN DE REGIONES PROMOTORAS DE GENES DE RIESGO A ESQUIZOFRENIA.	229
5.4.	ESTUDIO IV. ANÁLISIS DE TRANSCRIPTOMA DE GIRO CINGULADO EN ESQUIZOFRENIA. ESTUDIO DE VARIACIÓN GENÉTICA Y DIFERENCIAS DE EXPRESIÓN EN CEREBRO	245
	5.4.1. Discusión metodológica	245
	5.4.2. Discusión del análisis de expresión	248
	5.4.3. Discusión del análisis de variantes genéticas	255
6.	CONCLUSIONES	265
7.	BIBLIOGRAFÍA	269

ANEXO. Se adjunta un disco compacto con material suplementario al que se hace referencia a lo largo del proyecto de tesis presentado.





1. INTRODUCCIÓN



1.1. ASPECTOS CLÍNICOS Y NEUROBIOLÓGICOS DE LA ESQUIZOFRENIA

1.1.1. Antecedentes históricos

La esquizofrenia es la enfermedad mental por antonomasia. Como paradigma de la locura tradicional, ha sido objeto de estudio a lo largo de la historia y sufrido grandes cambios en las interpretaciones que se le fue dando con el paso de los años y los acontecimientos. Siempre representada como la forma de locura humana, desde la antigua Grecia se pueden encontrar diferentes descripciones e interpretaciones acerca de esta enfermedad.

Durante la edad media y la ilustración, la mala consideración que tenía el trastorno y el hecho de que siempre estuviera asociado a una degeneración personal fruto de condicionantes psicosociales fueron las causas de una serie de escenarios en general desfavorables para los enfermos. Por un lado, en zonas urbanas eran tratados en hospitales universitarios y, en zonas rurales, en asilos donde se trataba de paliar la sintomatología de forma caritativa. Los pioneros de la psiquiatría moderna trataron de cambiar esta práctica, pero no existían suficientes evidencias científicas para la legítima aplicación de una metodología similar a la elegida en otras áreas de la medicina (Heckers, 2011). En este contexto, también existía una minoría que, ya por aquel entonces, prefería abordar la enfermedad mental desde la exploración de los procesos subjetivos del inconsciente humano (Makari, 2008).

No fue hasta mediados del siglo XIX cuando se la empezó a considerar como a una enfermedad con una entidad propia. Concretamente, en 1850, el caso de un joven con un cuadro sintomático de deterioro mental prematuro le sirvió al médico francés Bénédict Morel para utilizar por primera vez el término *démence précoce* (Montes JM y Saiz J, 2000). No obstante, éste adquirió mayor repercusión tras la utilización del mismo por Emil Kraepelin (1856-1926) en lo que sería la sexta edición de su tratado de psiquiatría (Kraepelin, 1899). Su descripción de la *dementia praecox* agrupaba los síntomas de tres entidades, anteriormente descritas, pero que hasta el momento se encontraban muy desorganizadas: la catatonía, la hebefrenia y la demencia paranoide.

Esta última fue la primera aportación a la psiquiatría clínica de las tres mencionadas. Fue introducida por Wilhelm Sander (1838-1922) en 1868 para catalogar a un grupo de sujetos que tenían sentimientos infundados de maltrato, persecución o humillación constantes. Poco más tarde, en 1874, Kahlbaum (1828-1899) describió la catatonía (Goldar y cols., 1995) como un conjunto de síntomas bien reconocidos por los psiquiatras, aunque poco delimitados hasta entonces, entre los que destacaban el mutismo, la inmovilidad, la rigidez, el negativismo o la catalepsia. Por último, en 1871, Hecker (1843-1909) pormenorizó el concepto de hebefrenia

(Kraam y cols., 2009), explicándolo como un síndrome caracterizado por una desorganización en el comportamiento, en el habla, unas respuestas emocionales extremas, y un manierismo frecuente. De este modo, y retomando el concepto de demencia precoz, Kraepelin senta las bases de la futura esquizofrenia como una enfermedad sindrómica en la que factores biológicos individuales y ambientales pueden actuar conjuntamente y así provocar una degeneración temprana de las funciones cerebrales, manifestándose en tres "vertientes" del trastorno que comparten un diagnóstico común (demencia paranoide, catatonía y hebefrenia).

Asimismo, y dado que la *dementia praecox* aludía a la precocidad de la aparición y cronicidad de la enfermedad como características principales, se establecía una clara diferencia frente a otros síndromes con deterioro cognitivo de aparición tardía como la enfermedad de Alzheimer y, sobre todo, el grupo de las psicosis maníaco-depresivas (Montes y Saiz, 2000).

El gran impacto y enorme difusión en el campo de la psiquiatría que tuvo la obra de Kraepelin prepararon el camino para la aparición del concepto de esquizofrenia de Eugen Bleuler (Heckers y cols., 2011). Ampliamente influido por Kraepelin, Bleuler (1857-1939) acuñó el nombre actualmente conocido de Esquizofrenia, aunque por entonces se utilizaba el término *schizophrenias*, en plural, recalando la diversidad fenotípica dentro del mismo síndrome descrito. Bleuler, sin embargo, sugirió que el anterior término era inapropiado ya que el trastorno no era una demencia, es decir, no conllevaba necesariamente un deterioro de las funciones mentales.

Las nuevas ideas acerca de la Esquizofrenia tuvieron una gran influencia en Europa y en Estados Unidos. A diferencia de Kraepelin, Bleuler enfatizó en la descripción transversal de los síntomas frente a la menor importancia del curso y evolución de la enfermedad, y consideraba, además, que el trastorno del pensamiento era lo primario en la esquizofrenia y surgía como consecuencia de la escisión del proceso de formación de ideas y la respuesta emocional, lo que producía una fragmentación de las capacidades mentales. En este sentido, definió la sintomatología fundamental de la esquizofrenia con "*las cuatro Aes*": Asociaciones anormales de las ideas, Autismo, Afectividad inadecuada y Ambivalencia, mientras que los otros síntomas, como delirios y alucinaciones, serían accesorios e inconstantes. Además, a los tres subgrupos descritos por Kraepelin añadió uno nuevo, al que denominó esquizofrenia simple.

Desde entonces, el concepto de esquizofrenia se fue enriqueciendo a lo largo del siglo XX. Ejemplos de ello son las aportaciones de Kasanin, quien introdujo en 1933 el término *esquizoafectivo* para designar a un tipo de esquizofrenia con alteraciones afectivas importantes al margen del común trastorno formal del pensamiento (Kasanin, 1933), y la de Langfeldt, en 1939, que distinguió entre proceso esquizofrénico, de inicio temprano y deterioro severo, y reacción esquizofrénica, menos severo y capaz de remitir parcial o totalmente (Langfeldt, 1939).

Sin embargo, durante el siglo XX, y sobre todo tras la segunda guerra mundial, las ideas de Sigmund Freud impulsaron con gran fuerza el psicoanálisis como alternativa al modelo tradicional. Este método consideraba inútil el diagnóstico y la clasificación, y enfatizaba el esfuerzo en la naturaleza y el origen del *conflicto intrapsíquico* de cada caso. Dada la gran acogida en Estados Unidos en comparación con el resto del mundo, se creó una gran variedad diagnóstica que frenó el desarrollo de unos criterios objetivos para las enfermedades mentales (Andreasen, 2007). En este sentido, en Europa tuvieron gran impacto las teorías de Jacques Lacan, quien retoma la teoría psicoanalítica de Freud y le incorpora elementos del estructuralismo (Marta, 1987).

Además, el por entonces modelo de Bleuler carecía de unos criterios adecuados de pronóstico, y de unos rasgos propios y delimitadores de la enfermedad, lo que contribuyó a la ambigüedad en el concepto de la enfermedad y a la diversidad diagnóstica en los distintos países. Kurt Schneider (1887-1967), trató de solucionar ese problema postulando lo que él denominó síntomas de primer orden, y que, aunque no eran totalmente específicos de esquizofrenia, sí eran muy fáciles de identificar, muy fiables y cada uno de ellos una condición suficiente para realizar el diagnóstico de esquizofrenia. Entre ellos se encontrarían los delirios, fuga de pensamientos, voces constantes, pensamientos audibles o comportamientos extremadamente pasivos (Montes JM, y Saiz J, 2000). Muchos de estos síntomas se han adoptado actualmente como diagnóstico en las clasificaciones internacionales modernas. Finalmente, en la segunda mitad del siglo XX, la aparición de fármacos neurolépticos puso el foco de atención en la química del cerebro, impulsando la investigación biológica de la enfermedad y frenando el auge, casi de forma irreversible, de la alternativa psicoanalítica.

Sin embargo, todas estas aportaciones no sirvieron para delimitar de forma objetiva los límites diagnósticos de las enfermedades mentales. A raíz de un estudio realizado en 1972 (Cooper y cols., 1972) se observó que un mismo paciente era diagnosticado de esquizofrenia o psicosis maníaco depresiva según se encontrara en Londres o Nueva York, lo que puso de manifiesto la necesidad de utilizar unos criterios diagnósticos comunes. Surgen así los criterios de St. Louis (Feighner y col., 1972), que servirían de base para afianzar los criterios del manual diagnóstico y estadístico de los trastornos mentales (DSM) y de la Clasificación Internacional de Enfermedades (CIE), manuales que deben su existencia a la necesidad de unificar criterios a escala internacional.

A partir de ahí las investigaciones han seguido avanzando y, recientemente, ha sido publicado el DSM-5 donde se introducen algunos cambios muy significativos en el concepto y tratamiento de las enfermedades mentales, aunque leves en lo que a esquizofrenia respecta, tal y como se explica más detalladamente en el siguiente apartado.

1.1.2. Esquizofrenia. Definición y sintomatología.

La perspectiva histórica de la esquizofrenia, y el continuo desarrollo de su caracterización y delimitación sintomatológica a lo largo del siglo XX, nos revela su complejidad intrínseca. Como característica fundamental para entender el reto que supone su diagnóstico, es necesario señalar que, a día de hoy, no existe ningún marcador biológico que permita diagnosticar de forma objetiva a los pacientes, ni una causa única a la que atribuir el origen de la enfermedad. Por ello, la definición de esquizofrenia ha evolucionado a lo largo de los años a medida que se han ido aportando nuevas evidencias científicas y experiencias clínicas, ayudando a mejorar el diagnóstico y, en última instancia, el conocimiento global acerca de este trastorno psiquiátrico.

Con la reciente publicación del último manual diagnóstico de enfermedades mentales (DSM-5), son ya siete las ediciones del DSM (DSM-I, DSM-II, DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR y DSM-5) y 10 las del ICD (del ICD-1 al ICD-10), que han servido para establecer el concepto de esquizofrenia, definiéndola como una patología grave de tipo psicótico que generalmente presenta un carácter crónico, y engloba síntomas variados entre los que predominan una alteración de la afectividad, del comportamiento social, de la percepción de la realidad, del flujo y asociación de los pensamientos y la aparición de brotes psicóticos.

Por tanto, hablar de esquizofrenia en la actualidad es hablar de un síndrome complejo y muy heterogéneo en sus manifestaciones y síntomas, entre los que destacan la cronicidad y mal pronóstico (características claves en la noción de la *dementia praecox* de Kraepelin), síntomas negativos como la abulia, apatía, anhedonia o aplanamiento afectivo (fundamentales en la descripción de las esquizofrenias de Bleuler) y, por último, los síntomas positivos de Schneider (Schneider, 1959), que constituyen el fundamento de la “psicosis” e implican la presencia de ideas delirantes y alucinaciones manifiestas sin que la persona tenga conciencia de la naturaleza patológica de las mismas.

Durante el desarrollo de los manuales diagnósticos a lo largo de los años, los síntomas psicóticos de primer orden de Schneider y la cronicidad fueron ganando peso como condiciones necesarias para el diagnóstico de la enfermedad, y las posibles combinaciones de síntomas fueron constituyendo los diferentes tipos de esquizofrenia, en un esfuerzo por acotar cada variante de la enfermedad a unas manifestaciones concretas. De este modo, el DSM-4 reconoce cinco tipos de esquizofrenia: paranoide, desorganizado, catatónico, indiferenciado y residual, y el ICD-10 contempla nueve: paranoide, hebefrénica, catatónica, indiferenciada, post-esquizofrenia, residual, simple y las denominadas otras esquizofrenias y esquizofrenia no especificada.

No obstante, en los últimos años, varios estudios han demostrado que la distinción de subtipos de esquizofrenia resulta inconsistente, al no poderse distinguir con diferentes métodos de análisis clúster (Lykourasy cols., 2001; Peralta y Cuesta, 2003; Picardi y cols., 2012). Dada la diversidad sintomática de cada paciente y que estos presenten variaciones en los síntomas manifestados a lo largo del transcurso de la enfermedad, resulta muy complicado asignar a cada paciente una categoría estanca, y parece más evidente afirmar que un mismo sujeto puede pasar por varios de dichos subtipos de esquizofrenia a lo largo de la enfermedad. Una revisión de 24 publicaciones sobre 38 análisis de 28 cohortes diferentes reveló que no existe fundamento significativo para la clasificación tradicional en subtipos diferenciados (Linscott y cols., 2010). Además, menos del 5% de todas las publicaciones científicas distinguen entre sujetos provenientes de diferentes clases de esquizofrenia (Braff y cols., 2013).

Por otro lado, en la práctica clínica no existen mayores argumentos a favor de la utilización de los subtipos. A excepción de la indiferenciada y la paranoide, el resto de tipos raramente es utilizado. En un estudio chino de 19000 pacientes psiquiátricos hospitalizados a lo largo de 10 años, menos del 0,2% de los pacientes fueron diagnosticados como esquizofrenia catatónica, en el 1% fueron subtipo desorganizado y hasta el 91% recibió el diagnóstico de indiferenciado (Xu, 2011b).

Por estas razones, el reciente DSM-5 y el futuro ICD-11 (Braff y cols., 2013), eliminan las clases de esquizofrenia tal y como estaban descritas en anteriores versiones de los manuales. De este modo, se busca una mayor coherencia entre DSM e ICD, uno de los objetivos perseguidos en la nueva edición del DSM-5.

Sin embargo, pese a la eliminación de los tipos, en el actual DSM-5 no se producen cambios muy significativos en lo que al diagnóstico de esquizofrenia se refiere (Tandon y cols., 2013), y que se representa en la **tabla 1.1**. Esto es consecuencia del éxito cosechado en este punto en el DSM-4. Cerca del 90% del diagnóstico de esquizofrenia basado en éste se mantiene durante los siguientes 10 años (Bromet y cols., 2011). Por ello, solo se simplifica el criterio A, eliminando la excepcionalidad de los síntomas *schneiderianos* de rango I al no tener una implicación diagnóstica real (Cermolacce, Sass y Parnas., 2010) y estableciendo la necesidad de que al menos uno de los dos síntomas del criterio A sean alucinaciones ilusiones o habla desorganizada, y se restringe el trastorno esquizoafectivo a aquel que contenga sintomatología afectiva durante toda la enfermedad (Carpenter y Tandon., 2013).

Tabla 1. 1. Cambios en los criterios diagnósticos en el DSM-5 (Tabla extraída de Tandon y cols., 2013)

	Criterios DSM-4 en esquizofrenia	Criterios propuestos en DSM-5 para esquizofrenia
Criterio A. Síntomas característicos	<p>Dos o más de los siguientes en un período significativos de al menos un mes (o menos, si es tratado con éxito).</p> <p>1) Ilusiones 2) Alucinaciones 3) Habla desorganizada 4) Comportamiento desorganizado o catatonia 5) Síntomas negativos (p. ej, aplanamiento afectivo, alogia o abulia)</p> <p>Nota: Solo se requiere un síntoma del criterio A si las ilusiones son extrañas o consisten en una voz manteniendo un continuo comentario en los pensamientos o comportamientos del sujeto, o dos o más voces conversando entre ellas.</p>	<p>Dos o más de los siguientes en un período significativos de al menos un mes (o menos, si es tratado con éxito).</p> <p>Al menos uno de ellos debe ser de entre los síntomas 1-3</p> <p>1) Ilusiones 2) Alucinaciones 3) Habla desorganizada 4) Comportamiento desorganizado o catatonia 5) Síntomas negativos (p. ej, expresividad emocional reducida o abulia)</p> <p>Nota: Eliminada</p>
Criterio B. Disfunción social/ ocupacional	Para una significativa cantidad de tiempo desde el inicio del trastorno, una o más áreas del funcionamiento social como trabajo, relaciones interpersonales o auto cuidado, están marcadamente por debajo de los niveles previos al inicio (o, si es en la infancia o la adolescencia, incapacidad para alcanzar el nivel esperado en relaciones interpersonales y éxito académico u ocupacional).	Ningún cambio
Criterio C. Duración	Signos continuos del trastorno persistentes hasta, al menos, 6 meses. Este período debe incluir como mínimo un mes de síntomas (o menos si es tratado con éxito) del criterio A y puede incluir períodos de síntomas prodrómicos o residuales. Durante estos, los signos del trastorno se pueden ser manifestados solo por síntomas negativos o por dos o más de los síntomas del criterio A, presentes en la forma atenuada (p. ej, experiencias perceptivas inusuales o creencias extrañas).	Ningún cambio
Criterio D. Exclusión de trastorno depresivo y esquizoafectivo	Trastorno esquizoafectivo y depresión o trastorno bipolar con características psicóticas se descartan porque 1) No hay episodios depresivos mayores ni maníacos que hayan ocurrido de forma concurrente con la fase activa de los síntomas , o 2) si ocurren estos episodios durante la fase activa, su duración es breve en relación a la duración de la fase activa o síntomas residuales	Ningún cambio
Criterio E. Exclusión de trastorno por abuso de sustancias o condición médica	El trastorno no es atribuible al efecto de ninguna sustancia (p. ej, droga de abuso o medicación) o a una condición médica general	Ningún cambio
Criterio F. Relación con trastorno del desarrollo o del espectro autista	Si existe historia clínica de Autismo, el diagnóstico adicional de esquizofrenia se hace solo cuando si ilusiones o alucinaciones estás presentes al menos un mes (o menos si es tratado con éxito).	Si existe historia clínica de Autismo u otro trastorno comunicativo de inicio en la infancia , el diagnóstico adicional de esquizofrenia se hace solo cuando si ilusiones o alucinaciones estás presentes al menos un mes (o menos si es tratado con éxito).

No obstante, en cuanto a la descriptiva del diagnóstico, sí hay cambios significativos que alteran la noción previa de esquizofrenia que se extrae de anteriores versiones del manual. La catatonía, ya no un subtipo diferenciado, se incluye ahora como posible añadido a especificar (Carpenter y Tandon., 2013). También se profundiza más en los indicadores del curso de la enfermedad, alcanzando un mayor detallismo que en el DSM anterior. Por último y más importante, aunque no cobrando la relevancia diagnóstica que en principio parecía que iban a tener en el nuevo manual, se incluyen las dimensiones psicopatológicas de la enfermedad (Allardyce y cols., 2007; Hudziak y cols., 2009). La idea principal que subyace a la consideración de estas nuevas dimensiones es la sustitución de los subtipos por dominios de sintomatología en los cuales poder cuantificar la gravedad de cada uno de ellos (en una escala del uno al cuatro) a través del curso de la enfermedad. En el caso de esquizofrenia, estos dominios serían las cinco características presentes en el criterio diagnóstico A (ideas delirantes, alucinaciones, habla y/o comportamientos desorganizados, comportamiento psicomotor anormal y síntomas negativos) y otras tres relacionadas (deterioro cognitivo, manía y depresión). De este modo, se pretende abarcar el problema de la heterogeneidad de la esquizofrenia sin distinguir un síntoma por encima de los demás ni menospreciar otros que aparezcan aunque en menor grado (**figura 1.1**), dando al psiquiatra la oportunidad de generar una descripción multidimensional del paciente, a partir de cinco variables fácilmente identificables, que sea más informativa y que permita controlar las variaciones a lo largo de los años, sin la necesidad de incluir al sujeto en un subtipo que, en última instancia, no representa la complejidad de la esquizofrenia (Heckers y cols., 2013). La figura 1.1 refleja dos gráficos en donde se identifican rápidamente los rasgos distintivos de sendos ejemplos de esquizofrenia y trastorno esquizoafectivo sin ignorar los rasgos menos prominentes en cada uno de ellos y permitiendo evaluar el solapamiento.

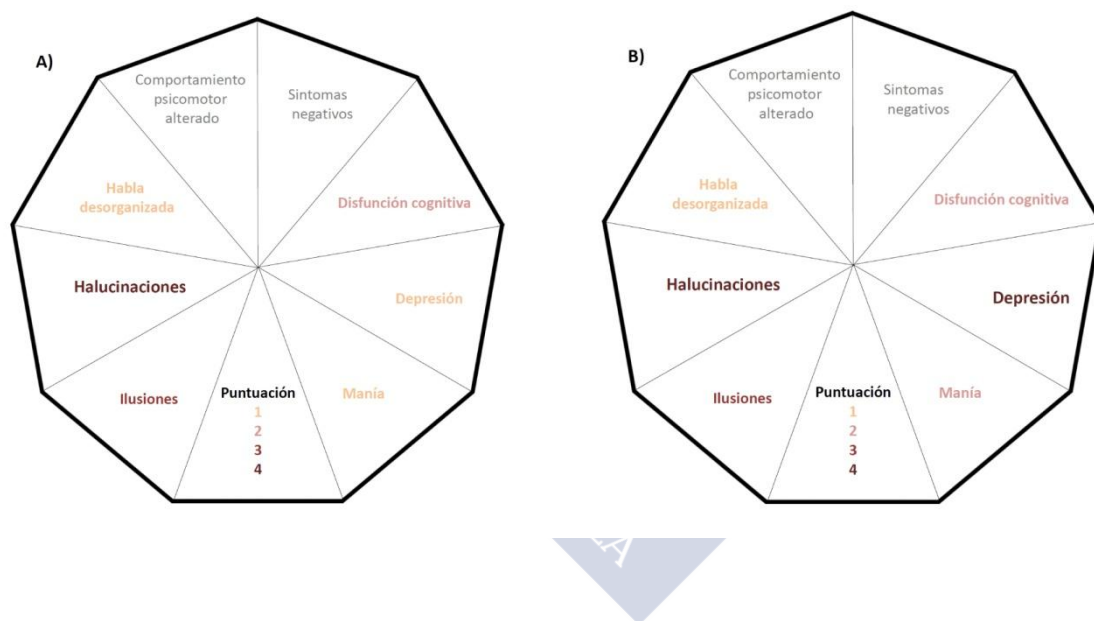
La propuesta inicial de integrar dimensiones en la práctica clínica puede sentar las bases para una futura concordancia entre la nosología psiquiátrica y la neurociencia clínica, facilitando el entendimiento de la enfermedad psiquiátrica y, por ende, de la esquizofrenia.

1.1.3. Curso de la enfermedad

De forma general, la esquizofrenia comienza durante la adolescencia o durante la vida adulta temprana, aunque no es un patrón repetido en todos los casos, y el curso de la enfermedad puede variar ostensiblemente. La esquizofrenia se puede manifestar como una repentina y rápida aparición o mediante un inicio más insidioso y subrepticio, y puede conllevar un único episodio o múltiples con diferentes grados de remisión en cada caso. Sin embargo, es bastante común la existencia de una fase prodrómica o pre-psicótica durante la infancia en la

cual los primeros indicios aparecen, fundamentalmente, en forma de síntomas negativos, aunque no se reciba atención clínica o tratamiento en la mayoría de los casos (Häfner y cols., 1999; Yung y McGorry; 1996). Este pródromo puede comenzar como cambios subagudos en el comportamiento (aislamiento social, comportamiento extraño o antisocial, deterioro de la capacidad académica o laboral), mal funcionamiento de las capacidades cognitivas, aparición de síntomas negativos progresivamente (alteraciones del habla, por ejemplo, habla incoherente o poco frecuente, o abulia), pensamientos desordenados (abstracto, paranoico y / o ideas obsesivas que pueden ser de naturaleza sexual, agresiva o irracional), despersonalización o desrealización (la percepción de ser ajeno a la identidad propia o con la realidad) entre otros (McGlashan., 1996).

Figura 1. 1. Evaluación dimensional de pacientes con A) Esquizofrenia y B) trastorno esquizoafectivo (figura adaptada de Heckers y cols., 2013)



Más recientemente, se ha empezado a considerar la aparición de la enfermedad como un proceso por etapas. Un modelo de desarrollo de la enfermedad aceptado por muchos autores es el que describe el proceso como cuatro etapas diferenciadas (Lieberman, 2001). La primera es una etapa premórbida inicial que va desde la gestación hasta la infancia, e incluso la adolescencia. Se manifiesta clínicamente por anomalías físicas, déficits cognitivos y una fuerte disfunción en la interacción social. Le sigue una etapa prodrómica, ya en la adolescencia, donde tiene lugar el episodio pre-psicótico descrito anteriormente. El siguiente paso sería la fase de inicio, que sucede con un episodio agudo psicótico en aproximadamente 50% de los casos (Ciompi, 1980). Por último, una posible etapa de deterioro crónico, en la cual deficiencias cognitivas y sintomatología psicótica pueden prolongarse hasta la senescencia. Un estudio

reciente reveló que son cuatro las variables más significativas a la hora de predecir la cronicidad de la esquizofrenia a raíz del primer episodio psicótico (Friis y cols., 2015): deterioro en el funcionamiento social, duración de la psicosis no tratada de más de 26 semanas, no remisión del episodio en los primeros tres meses y diagnóstico inicial de trastorno del espectro esquizofrénico.

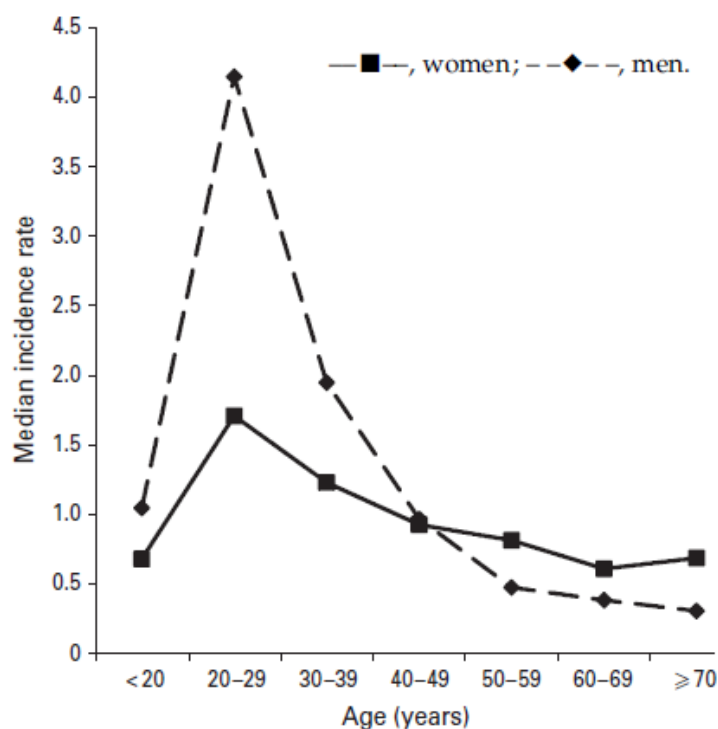
No obstante, existen diferencias importantes en la aparición y desarrollo de la enfermedad en función del sexo. Un reciente estudio que incluye 133693 casos de esquizofrenia de bases de datos variadas reveló que el riesgo en hombres es 1,15 veces mayor que en mujeres (van der Werf y cols., 2014) y que, aunque en promedio, la edad de aparición de la enfermedad es menor en hombres que en mujeres, lo es debido a la mayor intensidad en el inicio de la enfermedad en el pico de temprana adultez (**Figura 1.2**), siendo, al contrario, más frecuente en mujeres a partir de los 39 años. Ambos picos de iniciación de la enfermedad, en hombres y mujeres, se sitúan entre los 20 y 29 años. Estas evidencias contradicen los resultados de estudios previos, donde se describía una edad de inicio anterior en hombres y más cercana a la adolescencia (Häfner., 2003; Picchioni y Murray, 2007) y un segundo pico coincidente con la aparición de la menopausia, algo que ahora es imperceptible (**Figura 1.2**). Por otra parte, los síntomas en hombres son a su vez más severos que los acontecidos en mujeres (Leung y Chue, 2000).

Esta diferencia entre hombres y mujeres en la prevalencia y el transcurso de la enfermedad pone de manifiesto la dependencia de factores ligados al sexo como condicionantes de riesgo en esquizofrenia. A raíz de estas observaciones, diversas hipótesis se han planteado al respecto de este comportamiento. Entre otras, el desarrollo prenatal y la exposición diferencial a las hormonas gonadales en los hombres (Raz y cols., 1994), el retraso en la aparición de la enfermedad en mujeres debida a factores neuroprotectores como el estrógeno (Seeman, 1996), la mayor prevalencia de complicaciones en el parto en hombres (Clarke & O'Callaghan, 2003) o factores de riesgo social como la mayor probabilidad de contraer matrimonio en mujeres como factor protector (Morgan y cols., 2008).

Desde la eclosión inicial, existirá una alternancia entre fases agudas y fases de estabilización o post-crisis, dando cabida también a períodos atenuados o de mantenimiento, donde pueden llegar a desaparecer los síntomas casi por completo. La evolución de la enfermedad en cada paciente es muy variable, y mientras una proporción de enfermos recuperan la estabilidad tras el primer brote psicótico y no vuelven a sufrir ninguna recaída, otros sufren nuevos episodios, más o menos intensos y espaciados entre ellos, pudiendo llegar al peor pronóstico de estado crónico de la enfermedad. Aunque existe un rango muy amplio entre los diversos estudios al respecto, se estima que sólo un 30% de los pacientes evoluciona hacia un grado aceptable de remisión (Lambert y cols., 2010), habiendo un alto grado de ineficacia en el

tratamiento. En los casos crónicos, el grado de incapacidad es máximo y los perjuicios de la enfermedad se extienden a familiares y allegados, convirtiéndose el paciente en una carga enorme (Walker y cols., 2004), tanto emocional, por obvias razones de reactividad social, como económica, dada la incapacidad para llevar a cabo un trabajo dignamente y el consecuente alto desempleo entre los esquizofrénicos (Evensen y cols., 2015).

Figura 1. 2. Edad estimada e incidencia media de esquizofrenia en hombres y mujeres (Extraída de van der Werf y cols., 2014)



1.1.4. Epidemiología

Desde el punto de vista epidemiológico, el estudio de la esquizofrenia presenta muchos inconvenientes. Cabe destacar, entre ellos, la naturaleza multicausal del trastorno y su complejo diagnóstico, lo que podría provocar disparidad en los índices de prevalencia por factores de confusión ajenos a la propia variación intrínseca de la enfermedad. No obstante, los últimos estudios al respecto han corregido esta tendencia (Saha y cols., 2005)

La esquizofrenia es una enfermedad ampliamente distribuida sin excepción por todo el mundo. Sin embargo, pese a que el valor de prevalencia más citado cuando se habla de esquizofrenia es del 1%, los últimos estudios demuestran que se trata de un valor sobreestimado (McGrath, 2005), situándose la prevalencia a lo largo de la vida en un 0,4% y la prevalencia

puntual entre un 0,46% y un 0,5% a nivel mundial (Saha y cols., 2005; Messias y cols., 2007). Existen notables variaciones a nivel geográfico, como es el hecho de que las latitudes más altas tengan un factor de predisposición más alto (Saha y cols., 2006), relacionado con el hecho de haber nacido en invierno o primavera, que también ha sido demostrado como un factor de riesgo. Las diferencias pueden llegar a ser relativamente grandes como en la comparación de las prevalencias en Nueva Zelanda, de un 0,4%, y en Finlandia, de un 2,2%. Entre otros factores, también figuran variaciones debidas a diferencias culturales, regionales o étnicas. En una revisión sobre la incidencia en España, los datos se sitúan en un 0,3% en hombres y un 0,28% en mujeres (Ayuso-Mateos y cols., 2006). Estas diferencias en función del sexo no son exclusivas de España, sino que es un hecho global (van der Werf y cols., 2014), y es un factor de variación que ya se explica en detalle en el capítulo anterior, dejando claro que la esquizofrenia afecta en modo desigual a hombres y a mujeres.

La tasa de mortalidad estandarizada (TE) es una medida de la mortalidad relativa que se calcula mediante la división del índice de mortalidad observada en un período de tiempo de población bajo una determinada condición (esquizofrenia) con el esperado para esa población en ese período y con los parámetros de edad y sexo. La TE en esquizofrenia englobando todas las causas existentes a finales del siglo pasado se situaba en un 1,5 (Brown, 1997), pero este valor crece de forma preocupante con los años debido a la menor tasa de mortalidad en la población general (McGrath y Sasser, 2009) y a día de hoy se sitúa en 2,68. La TE puede dividirse en causas concretas. En una revisión reciente se recogen datos de mortalidad y se calculan las TE para 561 posibles causas (Saha y cols., 2007). El caso del suicidio es de mayor valor, siendo de una TE de 12,6. Sin embargo, no es nada desdeñable el hecho de que la mayor parte de las causas de muerte estén elevadas en esquizofrenia, pese a que *a priori* pareciera que poco o nada tienen nada que con el trastorno, como enfermedades endocrinas, digestivas o respiratorias.

La exposición prenatal a infecciones por parte de la madre es un factor de riesgo contrastado para el desarrollo de la esquizofrenia (Buka y cols., 2001). Existen evidencias concretas que sugieren que el riesgo a sufrir esquizofrenia es mayor en aquellos con la exposición prenatal al virus de la gripe o *influenza* (Brown y cols., 2004a), *rubeola* (Brown y cols., 2014) o *Toxoplasma gondii* (Brown y cols., 2005; Mortensen y cols., 2007). La nutrición también se ha postulado como otro factor de riesgo, siendo algunos de los elementos propuestos la deficiencia a vitamina D (McGrath y cols., 2003) o la presencia excesiva de homocisteína durante el embarazo en el útero materno, y su repercusión en el metabolismo del folato (Brown y cols., 2007). Las conocidas hambrunas holandesa y China, debidas a períodos de desnutrición causados por la invasión nazi durante la segunda guerra mundial y por la revolución cultural, respectivamente, revelaron un mayor índice de esquizofrenia entre la progenie de madres en

riesgo de desnutrición (Susser y cols., 1992; St Clair y cols., 2005), pudiendo haber mediado, en estos casos, mecanismos epigenéticos.

Las complicaciones obstétricas y del embarazo son otro gran grupo de factores de riesgo entre los que podríamos destacar nacimiento prematuro, preeclampsia, diabetes gestacional o malformaciones congénitas (Cannon, 2002), habiendo una asociación modesta pero consistente, con un ratio de 2 respecto al riesgo de la población general.

También se ha demostrado que la edad paterna es un factor de riesgo importante a esquizofrenia y a otras enfermedades psiquiátricas (Malaspina y cols., 2015). Una posible explicación a este hecho sería la presencia de mutaciones *de novo* en genes importantes en el desarrollo neuropsiquiátrico en la línea germinal paterna, así como posibles alteraciones epigenéticas que afecten a los mismos con la edad, como fallos en el empaquetamiento de la cromatina o cambios en las modificaciones químicas de las histonas. Sin embargo, otros estudios señalan que el hecho de que la edad paterna avanzada no sería la causa *per se*, sino que otros factores asociados a una paternidad retrasada serían los posibles causantes de esta asociación (Ek y cols., 2014).

La condición de emigrante o pertenencia a un grupo minoritario y/o clase social desfavorecida son también factores de riesgo, así como la residencia en entorno urbano, consumo de drogas y estado civil separado o soltero. Una interesante discusión al respecto es si estos factores sociológicos predisponen al desarrollo de esquizofrenia por alterar sus funciones cerebrales óptimas en combinación con la predisposición genética (van Os y cols., 2010) o bien si una alteración previa de ese *cerebro social* es lo que impulsa a estos individuos a acabar cayendo en estas situaciones que representan una consecuencia clara de la inadaptación social (Sullivan, 2005). La asociación con el consumo de drogas como el cannabis, demostrada en numerosas ocasiones (Murray y cols., 2007), ha sido objeto de discusión ya que la propia alteración de las funciones cerebrales podría predisponer a un consumo de cannabis, de forma que pudiéramos estar en el contexto de una asociación indirecta (Cassidy y cols., 2014).

Por último, otro aspecto clave en la esquizofrenia es la reducida fertilidad en los enfermos (Bundy y cols., 2011). Dada la incapacidad de los enfermos de establecer relaciones sociales estables y contraer pareja, el índice de fertilidad se reduce en un 39%, lo cual ha llevado a muchos autores a pensar que debe de existir algún mecanismo de selección positiva que confiera algunas ventajas a individuos con un trasfondo genético similar al de la esquizofrenia pero que no llegan a desarrollar la enfermedad; o bien que se trata de una consecuencia inherente a la evolución de la inteligencia humana (Crow, 1997; Crow, 2008).

1.1.5. Neuroanatomía y alteraciones estructurales

La esquizofrenia es un trastorno complejo con una sintomatología muy diversa y heterogénea. A lo largo de los últimos años, diversos estudios se han servido de técnicas variadas para el estudio de la enfermedad desde el punto de vista de la anatomía cerebral y de la funcionalidad del mismo. Sin embargo, durante la primera mitad del siglo XX, los esfuerzos para identificar anomalías en los cerebros *post mortem* llevaron a resultados decepcionantes por lo rudimentario de las técnicas y la gran expectación no satisfecha.

No fue hasta los años setenta, con la llegada de las técnicas de **neuroimagen estructural** como la Resonancia Magnética (RM) y la Tomografía Axial de Positrones (TAC) que se encontraron anomalías en los cerebros de los enfermos. Estas técnicas proporcionan una visión anatómica de las alteraciones estructurales en la esquizofrenia. Gracias a ellas se produjeron hallazgos muy importantes, como la reducción del volumen y del peso total del cerebro, sobre todo localizado en el lóbulo temporal medial y superior, hipocampo, corteza prefrontal y tálamo, observables desde la fase inicial de la enfermedad. Por otro lado, también se han encontrado incrementos de volumen en el putamen y la corteza estriada (Meyer-Lindenberg, 2010). No obstante, de los estudios de neuroimagen estructural se concluye que no existen lesiones como tales, pero que sí reducciones en el volumen y pérdida de densidad de ciertas regiones (Cuevas-Esteban y cols., 2011).

Estos cambios macroscópicos se han asociado a alteraciones a escala microscópica, como la reducción en el número de neuronas y ramificaciones dendríticas. Las neuronas piramidales en la corteza prefrontal, que constituyen la principal fuente de neurotransmisión, son más pequeñas y presentan un empaquetamiento más denso, debido probablemente a una excesiva poda sináptica durante la pubertad, esto es, una eliminación de las conexiones neuronales no utilizadas durante el desarrollo para que la capacidad cerebral este focalizada en las conexiones frecuentemente utilizadas (Meyer-Lindenberg, 2010).

A diferencia de los escasos incrementos en el volumen de regiones cerebrales, las reducciones en el hipocampo han sido halladas en familiares de pacientes y personas sanas, lo cual revela que debe tener una profunda base genética (Goldman y cols., 2008). Sin embargo, las alteraciones observadas no son privativas de enfermos, y muchos pacientes no presentan alguna de estas alteraciones. Se ha visto, además, que otros muchos factores pueden provocar alteraciones similares, como en el caso del alcohol (Squeglia y cols., 2014), la nicotina (Brody y cols., 2004) o el tratamiento con antipsicóticos (Veijola y cols., 2014), así que en ningún caso servirían como factores patológicos que asegurasen el diagnóstico de la enfermedad. A esto habría que añadirle el hecho de que, tal y como indican muchos estudios, la esquizofrenia es un

trastorno del neurodesarrollo, y los cambios estructurales ya se habrían ido gestando antes de la eclosión de los primeros síntomas diagnosticables (Owen y cols., 2011).

Por otra parte, los trabajos de **neuroimagen funcional** también fueron ganando peso a partir de los años setenta junto con el desarrollo de técnicas de imagen más sofisticadas que permitiesen la evaluación de anomalías cerebrales en funciones concretas como la atención, la memoria o alteraciones sensitivo-motoras, entre otras, proporcionando un acceso directo a la investigación de la fisiopatología de la esquizofrenia *in vivo* (Gur y cols., 2007). Técnicas como la Tomografía por emisión de positrones (TEP) o la Resonancia magnética funcional (RMf) han sido las responsables de los hallazgos encontrados, aunque la segunda ha ido ganando peso con los años en detrimento de la primera debido a que ésta es más invasiva y conlleva un coste económico mayor. Estos estudios de neuroimagen han puesto de relieve que hay una actividad y conectividad anormal en los circuitos cerebrales que comprenden el córtex prefrontal y el lóbulo temporal (Salgado-Pineda y cols., 2004), además de una alteración de las conexiones entre lóbulo frontal y temporal y otras áreas del cerebro (Meyer-Lindenberg, 2010).

De igual modo que con las anomalías estructurales, también en el caso de las funcionales sus hallazgos no se limitan a pacientes y tampoco son condición necesaria en todos ellos, confirmando la imposibilidad de ser la base de un diagnóstico fiable, pese a su innegable utilidad.

Al margen de los estudios de neuroimagen, se ha observado, también, la presencia de otros rasgos funcionales alterados en individuos esquizofrénicos, como déficits en la memoria verbal y a corto plazo, alteraciones en la estructura del sueño o la memoria prospectiva, relacionadas con la incapacidad de llevar a cabo una vida independiente (Ordemann y cols., 2014). Además, otra de las funciones más afectadas en la esquizofrenia es la que engloba a los procesos de percepción y, en especial a los auditivos. Es el caso de la inhibición en la amplitud de potenciales evocados ante estímulos repetidos o la inhibición del prepulso del reflejo de la respuesta de sobresalto ante un estímulo acústico brusco (PPI). La intensidad de dicha inhibición se ha establecido como una medida del procesamiento de la información sensorial y motora, y su ausencia parcial o total estaría vinculada a la incapacidad para filtrar los estímulos irrelevantes (Jazbec y cols., 2007), generando una sobreexcitación en los sistemas dopaminérgico y glutamatérgico y una fragmentación cognitiva. De nuevo, estos rasgos funcionales son, además, compartidos por familiares no afectados, calculándose una heredabilidad en torno al 50% (Greenwood y cols., 2007). Esto, al igual que ocurría con las alteraciones estructurales, viene a demostrar que estos son rasgos heredables y que la genética debería poder explicar su incidencia.

1.1.6. Neurobiología de la esquizofrenia

Conocer las bases biológicas que explican el inicio y desarrollo de la esquizofrenia es condición necesaria para el entendimiento de la enfermedad y la posible implementación de medidas preventivas. Sin embargo, dada la complejidad intrínseca de esta enfermedad, nuestro conocimiento de la misma, desde el punto de vista etiológico, todavía se limita a hallazgos experimentales que, en la mayoría de los casos, aportan relaciones de causalidad indirecta con la sintomatología de la enfermedad. Actualmente, dichas evidencias son resultado de los múltiples y diversos estudios que se han realizado a lo largo de los últimos años para determinar las alteraciones que pudieran existir en el cerebro de los pacientes con esquizofrenia desde el punto de vista anatómico, fisiológico o neuroquímico.

1.1.6.1. Hipótesis neuroquímicas.

Pese a que el mecanismo neuroquímico de la esquizofrenia permanece en parte desconocido, en la actualidad existe una variedad de hipótesis que implican la disfunción del normal funcionamiento de sistemas de neurotransmisores, en muchos casos de forma interrelacionada, en el desarrollo de la esquizofrenia (Saiz Ruiz y cols., 2010; Keshavan y cols., 2011).

La primera hipótesis formulada en este sentido, y con mayor aceptación inicial, fue la **hipótesis dopaminérgica**. Esta teoría, que explicaría la sintomatología positiva de la enfermedad, surgió al comprobarse, por serendipia, que los primeros antipsicóticos utilizados en pacientes esquizofrénicos eran antagonistas de los receptores de dopamina en el cerebro (Carlsson y Lindqvist., 1963), apoyando la posibilidad de que los sujetos enfermos estuviesen sobreexpuestos a este neurotransmisor, y dando lugar al modelo hiperdopaminérgico (Carlsson, 1977). De esta forma, los antipsicóticos bloquearían, en parte, la excesiva captación de dopamina de los receptores D2, eliminando la sobreexcitación sináptica y llevándola a unos niveles normales. Es el caso de fármacos como el *Haloperidol* o la *Clorpromazina*. Paralelamente a los efectos terapéuticos, se observaron en los pacientes una serie de efectos secundarios semejantes a los síntomas de la enfermedad de Parkinson (en la que existe una deficiencia de neuronas dopaminérgicas) como temblores, rigidez muscular y disminución general de los movimientos voluntarios. Por otro lado, también se observó que los agentes liberadores de dopamina producían efectos psicóticos (Thomson, Pogue-Geile y Grace., 2004). Tal es el caso de drogas de abuso de carácter psicotrópico, como la cocaína o las anfetaminas, que pueden provocar cuadros psicóticos similares a los de la esquizofrenia.

Actualmente, sin embargo, se contempla un escenario más complejo en la hipótesis que relaciona la dopamina con la esquizofrenia, planteándose un modelo de *desequilibrio dopaminérgico límbico – prefrontal* (Saiz Ruiz y cols., 2010; Weinberger, 1987). Por un lado, existiría una hiperfunción dopaminérgica subcortical, en la vía mesolímbica, que proyecta el neurotransmisor desde el mesencéfalo a ciertas áreas límbicas como el núcleo accumbens, que forma parte del circuito de recompensa. La hiperactividad en esta vía provocaría una hiperestimulación de los receptores de dopamina D2 y explicaría los cuadros de sintomatología positiva. Por otro lado, existiría también una hipoactividad en la vía mesocortical, que proyecta desde el área tegmental al córtex prefrontal, y que se relaciona con funciones como la regulación de emociones y afectividad. De esta forma, la hipoestimulación de los receptores de dopamina D1 explicaría la sintomatología negativa y déficits cognitivos propios de la esquizofrenia (Cropley y cols., 2006). Este modelo describe de forma integrada la coexistencia de síntomas positivos y negativos pero, sin embargo, no da una respuesta al deterioro funcional o a las alteraciones estructurales en los cerebros de los enfermos.

La alteración del sistema excitatorio de neurotransmisión del glutamato, y sobre todo del receptor de glutamato N-metil D-Aspartato (NMDA), constituye la **hipótesis glutamatérgica**, que ha cobrado mayor interés en los últimos años. Surgió tras el descubrimiento de que pacientes de esquizofrenia tenían niveles bajos de glutamato en fluido cerebrospinal (Kim y cols., 1980), y de que, bajo la acción de los antagonistas *fenciclidina*, *Ketamina* o *MK-801*, los receptores NMDA son hipofuncionantes, provocando la aparición de sintomatología positiva, negativa y cognitiva propia de la esquizofrenia (Olney y Farber, 1995). Esta hipótesis integraría, además, los posibles efectos de la desregulación dopaminérgica comentada anteriormente ya que la hipofunción del receptor NMDA en las proyecciones *corticoencefálicas* conllevaría una hipoactividad en la *vía* mesocortical dopaminérgica y una hiperactividad dopaminérgica mesolímbica, explicando de esta manera la aparición de síntomas negativos, afectivos y cognitivos, o síntomas psicóticos, respectivamente (Coyle y cols., 2006). Además, y contrariamente a lo ocurrido con la hipótesis dopaminérgica, la hipofunción del receptor del glutamato NMDA provocaría una hiperexcitabilidad glutamatérgica y podría explicar el proceso de deterioro crónico, dada la neurotoxicidad inducida por el glutamato, y que muchos de los hallazgos genéticos más importantes implican a genes de esta vía (Carter, 2006). Sin embargo, este modelo no aclara por qué, a pesar de la ubicuidad de neuronas glutamatérgicas en el cerebro, la esquizofrenia no se encuentra asociada a un deterioro neurológico más generalizado (Crow, 1995).

Pese a que las anteriores son las vías de mayor aceptación en relación a la esquizofrenia, hay evidencias importantes que apuntan a la implicación de otras vías de neurotransmisión. Una de ellas constituye la **hipótesis gabaérgica**. El sistema del neurotransmisor *ácido γ -*

aminobutírico (GABA), principal inhibidor en el sistema nervioso central (SNC) en mamíferos, se ha visto implicado en esquizofrenia (Lewis y Hashimoto, 2007). Estudios *post mortem* en pacientes han encontrado niveles reducidos de neuronas gabaérgicas en córtex prefrontal, así como de la descarboxilasa del ácido glutámico, una enzima clave en el proceso de síntesis de GABA (Lewis y cols., 2005).

La **hipótesis serotoninérgica**, por su parte, surgió con el estudio del mecanismo de actuación de los antipsicóticos de segunda generación como la *clozapina*, donde se observó que éstos interactuaban con los receptores de serotonina, al igual que ocurría con el ácido lisérgico y la *psilobicina*, agonistas del receptor de serotonina y causante de sintomatología positiva similar a la esquizofrénica (Geyer y Vollenweider, 2008). Los diversos estudios que implican a la serotonina destacan su rol en el neurodesarrollo, su interacción con el sistema dopaminérgico y los efectos de la serotonina en la corteza prefrontal a través de sus receptores *5HT2A* (Kapur y Remington, 1996).

Por último, y aunque en menor grado de relevancia, también se ha barajado una posible **hipótesis colinérgica**, donde una hiperactividad de este sistema de neurotransmisión pudiera estar de la sintomatología negativa de la enfermedad (Tandon y Greden, 1989). De hecho, la administración de agonistas colinérgicos provoca un incremento de los síntomas negativos en pacientes con esquizofrenia y una aparición de estos síntomas en individuos sanos (Peralta y Cuesta, 1995).

1.1.6.2. Hipótesis inflamatoria

En los últimos quince años, ha ido ganando peso la hipótesis de la respuesta inflamatoria como un potente factor de riesgo en la enfermedad mental y, particularmente, en la esquizofrenia (Kirkpatrick y Miller, 2013; Miller y cols., 2013a). Numerosos estudios apoyan la idea de que una respuesta inmunológica o inflamatoria alterada, sobre todo en las etapas más tempranas de la vida, podría tener una fuerte influencia en el inicio y posterior desarrollo de la esquizofrenia y otros trastornos relacionados, explicando, en parte, la gran comorbilidad existente entre trastornos psiquiátricos y otras enfermedades donde el sistema inmune y la respuesta inflamatoria están directamente implicados como la diabetes, la intolerancia a la glucosa o la hipertensión (Bushe y Holt, 2004; Friedman y cols., 2010).

La inflamación es la respuesta defensiva del organismo a la lesión o infección y puede presentarse en forma de episodios agudos, caracterizados por acciones fisiológicas más agresivas y notorias como fiebre, calentamiento, dolor, sudoración, permeabilización de vasos sanguíneos circundantes y migración de leucocitos, o en forma de inflamación crónica, más

silenciosa y latente, y característica de muchas patologías como Alzheimer, diabetes u obesidad crónica. Las citoquinas son elementos claves en este proceso. Consideradas las hormonas del sistema inmune, engloban proteínas de diferente naturaleza química (interferones, interleuquinas, factores de necrosis tumoral,...) y regulan el proceso de un modo pro-inflamatorio (*IL-1*, *IL-6*, *IL-8*, *TNF α* o *IFN γ*) o antiinflamatorio (*IL-1RA*, *IL-4* o *IL-10*), guardando un equilibrio entre ambas funciones para optimizar la respuesta. Secretadas en todo el organismo por diversos tipos celulares, son el principal vehículo de información entre el sistema inmune y el SNC en el cual su principal fuente son las células de la microglía. La activación de este tipo celular provoca la respuesta inflamatoria mediada por citoquinas que regulan el proceso inflamatorio y, en última instancia, tienen la capacidad de atravesar la barrera hematoencefálica, actuando también en el interior cerebral. Las hipótesis que vinculan la respuesta inflamatoria con la esquizofrenia se basan principalmente en los numerosos estudios que reportan alteraciones en la concentración de citoquinas en pacientes esquizofrénicos e individuos con síntomas prodrómicos. Estas observaciones, además, son extensibles a muchos otros trastornos neuropsiquiátricos, y parece una condición general de las enfermedades complejas que afectan al SNC.

La mayor evidencia de la implicación del sistema inmunitario en el desarrollo de la esquizofrenia procede de la exposición prenatal a la inflamación, principalmente vía materna. De esta forma, la inflamación podría ser un mediador de muchos factores de riesgo prenatales conocidos, como diabetes gestacional, estrés, depresión o preeclampsia (hipertensión inducida por el embarazo) (Miller y cols., 2013a), siendo esta última, de por sí, un factor de riesgo de aparición de esquizofrenia. En humanos, análisis serológicos de madres que dieron lugar a prole con esquizofrenia y psicosis, recogidas durante el segundo trimestre del embarazo, revelaron una concentración significativamente alta de *IL-8* (Brown y cols., 2004b) y de *TNF α* (Buka y cols., 2001), respectivamente. Además, otros estudios animales también sugieren que la inflamación durante períodos críticos del desarrollo neurológico puede provocar un aumento de la inflamación en los hijos adultos, mediante alteración de niveles de *IL-1 β* , *IL-1RA*, *IL-6* o *IL-10* (Boksa, 2010; Meyer y cols., 2009). En esta línea, otros estudios con tamaños muestrales más grandes demostraron un mayor riesgo a padecer psicosis en hijos cuyas madres habían sufrido algún tipo de infección (Zammit y cols., 2009) o que habían tomado medicación antiinflamatoria durante el embarazo (Sørensen y cols., 2004). Meyer, Feldon y Dammann (2011) proponen un interesante modelo en el cual esquizofrenia y autismo comparten un origen inflamatorio prenatal común, hasta que, llegado un punto del desarrollo, determinantes genéticos y ambientales dividen el destino de ambas enfermedades con una persistente o latente inflamación en la adultez de individuos con autismo o esquizofrenia, respectivamente.

Sin embargo, las evidencias inflamatorias no se limitan al estado prenatal. Individuos con esquizofrenia muestran niveles elevados de citoquinas proinflamatorias en sangre (Potvin y cols., 2008), hecho que ocurre, además, en pacientes en su primer episodio psicótico, lo cual descarta que el "perfil pro-inflamatorio" observado se deba al efecto de la medicación antipsicótica. Por otro lado, las concentraciones en sangre de algunas citoquinas y otros agentes inmunológicos podrían ser marcadores del estado de la enfermedad, como ocurre con el aumento de *IL-6* o de linfocitos detectado en fases agudas de sintomatología psicótica y en recaídas (Miller y cols., 2013b). En otros hipotéticos marcadores de rasgo del trastorno, como el caso de *IL-12* o *TNF α* , un aumento en la concentración de éstas se mantiene estable y se asocia a una mayor cronicidad en el trastorno (Miller y cols., 2011). Pese a que estos estudios relacionan la enfermedad con inflamación periférica, existen evidencias de que la alteración en niveles periféricos de citoquinas puede estar relacionada con otros trastornos que afectan al cerebro, tales como Alzheimer, deterioro cognitivo (Sardi y cols., 2011) o depresión (Dowlati y cols., 2010). Además, los niveles periféricos de citoquinas interaccionan directamente con los sistemas dopaminérgico y serotoninérgico (Song y cols., 1999) e indirectamente con el sistema glutamatérgico, mediante la inducción de la degradación de triptófano en ácido quinurénico, antagonista del receptor NMDA (Müller y Schwartz, 2008). Estos hallazgos, a la vez, proporcionan un interesante vínculo entre las hipótesis neuroquímicas anteriormente descritas, que implican alteraciones en los sistemas de neurotransmisores, y los niveles anómalos de citoquinas, producto de respuestas inflamatorias alteradas muy probablemente debido a factores de origen prenatal.

Aunque en menor número y debido a una mayor complicación metodológica, otros estudios también han encontrado niveles anormales de varios parámetros de respuesta inflamatoria en el SNC en esquizofrenia, como citoquinas y leucocitos en fluido cerebroespinal (Ganguli y Gubbi, 1997; Nikkilä y cols., 2001), linfocitos y activación de microglía (Busse y cols., 2012) o elevados niveles de estrés oxidativo (Do y cols., 2010).

Recientemente, Bloomfield y cols. (2015) publicaron un novedoso estudio en el cual demostraban la activación de microglía *in vivo* en individuos esquizofrénicos, individuos de alto riesgo a sufrir esquizofrenia y controles, mediante tomografía de emisión de positrones. La comparativa de los grupos tras corregir por los posibles factores de confusión reveló un aumento significativo de la actividad en aquellos individuos de alto riesgo comparable a los esquizofrénicos y perfectamente distinguible de los controles, lo cual podría tener muchas implicaciones en acciones de prevención y/o diagnóstico en la práctica clínica en un futuro cercano. Este hallazgo se suma a los otros estudios previos donde también se encontraron alteraciones neurofisiológicas o anatómicas propias de esquizofrenia en modelos de inflamación.

Como consecuencia de esta respuesta inmune, podría explicarse el incremento en el estrés oxidativo, entendido como la acumulación de especies de oxígeno reactivas, y que ya ha sido reportado en varios estudios (Bitanirwe y Woo, 2011; Prabakaran y cols., 2004; Wang y cols., 2009a), donde se observa una reducción en los procesos antioxidantes o aumento de los oxidantes, con el consiguiente daño neuronal que estas especies con radicales libres, y por ello muy reactivas, podrían provocar.

Además, desde el punto de vista genético, todo apunta a que una disfunción en genes asociados a la respuesta inmune y al proceso inflamatorio estaría directamente implicada en el desarrollo de esquizofrenia. En un estudio Finés, se observó que la prevalencia de infecciones en el tracto urinario es cinco veces mayor en aquellas mujeres embarazadas con historial familiar de psicosis (Clarke y cols., 2009). Evidencias moleculares preliminares en estudios de asociación o de *pathways* sugieren una implicación del sistema inmune en la enfermedad (Paul-Samojedny y cols., 2010, Hsu y cols., 2009). Otros estudios revelan, además, interacciones significativas entre genes candidatos conocidos y otras proteínas de respuesta inflamatoria, como es el caso de polimorfismos en *NRG1* y la *interleuquina IL-1* (Hänninen y cols., 2007) o la interacción entre *DTNBP1* y *IL-3* (Edwards y cols., 2008), y que variantes en genes de respuesta inmune también se asocian a alteraciones morfológicas propias del trastorno (Papiol y cols., 2005). Sin embargo, la principal evidencia es que la asociación genética más significativa en estudios de asociación de genoma completo (GWAS) implica a proteínas de la clase I del complejo mayor de histocompatibilidad (MHC) (Purcell y cols., 2009; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Shi y cols., 2009; Stefansson y cols., 2009), tal y como se desarrolla en mayor profundidad en siguientes apartados. A nivel de expresión, también se vieron niveles aumentados de ARN mensajero en pacientes esquizofrénicos (Saetre y cols., 2007). No obstante, dado el alto grado de desequilibrio de ligamiento (DL) en la región del MHC, y la existencia de genes no implicados en inmunidad en ella, no se puede descartar que estas asociaciones sean indirectas, y la causalidad implique realmente a otros genes.

1.1.6.3. Hipótesis del neurodesarrollo.

La hipótesis del neurodesarrollo como modelo explicativo de la esquizofrenia fue propuesto en primer lugar hace ya treinta años aproximadamente (Weinberger, 1986; Murray y Lewis, 1987). La idea principal que postulaba esta teoría era que la esquizofrenia, pese a manifestarse clínicamente en la edad adulta, debía su origen a perturbaciones ocurridas durante el desarrollo del sistema nervioso y el cerebro. Posteriormente, diversas evidencias como las anomalías en cerebros reveladas en estudios de neuroimagen, estudios epidemiológicos o

alteraciones cognitivas y motoras en individuos en los que posteriormente se observó el desarrollo de la enfermedad, dieron consistencia y credibilidad a este modelo. Otros trabajos demostraron que lesiones en primates neonatos podían acarrear desórdenes en el comportamiento análogos a los observados en esquizofrenia (Owen y cols., 2011; Rapoport y cols., 2012).

Durante la fase de **neurodesarrollo temprano**, acontecimientos ambientales o genéticos (como deleciones o duplicaciones de material genético) que tienen repercusión en procesos como la neurogénesis, proliferación, diferenciación o migración neuronal provocan alteraciones en la maduración del cerebro y su capacidad para interactuar con el entorno y su respuesta a factores de estrés ambiental posteriores (Ashe y cols., 2001), dando lugar a la eclosión de la enfermedad en un punto determinado de la adolescencia o adultez, con un posible factor desencadenante, pero con el trasfondo de neurodesarrollo como elemento de predisposición más determinante. Estos cambios serían claves en los procesos de maduración cerebral, sobretudo en áreas corticales (más implicadas en la esquizofrenia tal y como se ha visto anteriormente). De esta forma, la disfunción en la respuesta cerebral a situaciones de estrés se plasmaría en una hipersensibilidad a las mismas y en consecuencia, en una mayor susceptibilidad a sufrir sintomatología psicótica ante determinados estímulos, actuando como “puntos gatillo” que dispararían la aparición de la enfermedad (Lewis y Levitt, 2002).

No obstante, la aparición tardía de la enfermedad en contraste con las evidencias que vinculan la enfermedad a acontecimientos del desarrollo temprano del cerebro también podrían deberse a otros procesos del **neurodesarrollo tardío** ocurridos durante la adolescencia, como la mielinización, poda sináptica (para especializar funciones neuronales concretas) o apoptosis neuronal. De esta forma, muchos autores defienden que el origen de la enfermedad reside principalmente en las alteraciones neuroquímicas ocurridas durante estas etapas tardías, coincidentes con la aparición de la enfermedad. Sin embargo, a día de hoy existen muchas evidencias, tanto genéticas como ambientales, que demuestran que alteraciones en el neurodesarrollo temprano, como mínimo, contribuyen notablemente a la probabilidad de sufrir esquizofrenia (Rapoport y cols., 2012).

La explicación de la hipótesis del neurodesarrollo genera, además, un modelo que englobaría los rasgos fenotípicos encontrados y replicados en el estudio de la esquizofrenia, tanto anatómicos como neuroquímicos y funcionales. En particular, las alteraciones del sistema inmune y la respuesta inflamatoria, anteriormente descritas, tienen una mayor relevancia durante las etapas iniciales de neurodesarrollo, mucho antes de la aparición de las primeras manifestaciones clínicas. En este sentido, son ejemplos claves las implicaciones de infecciones

durante el embarazo o la asociación de la enfermedad con variaciones genéticas que puedan afectar a la regulación de genes expresados particularmente en etapas tempranas del desarrollo.

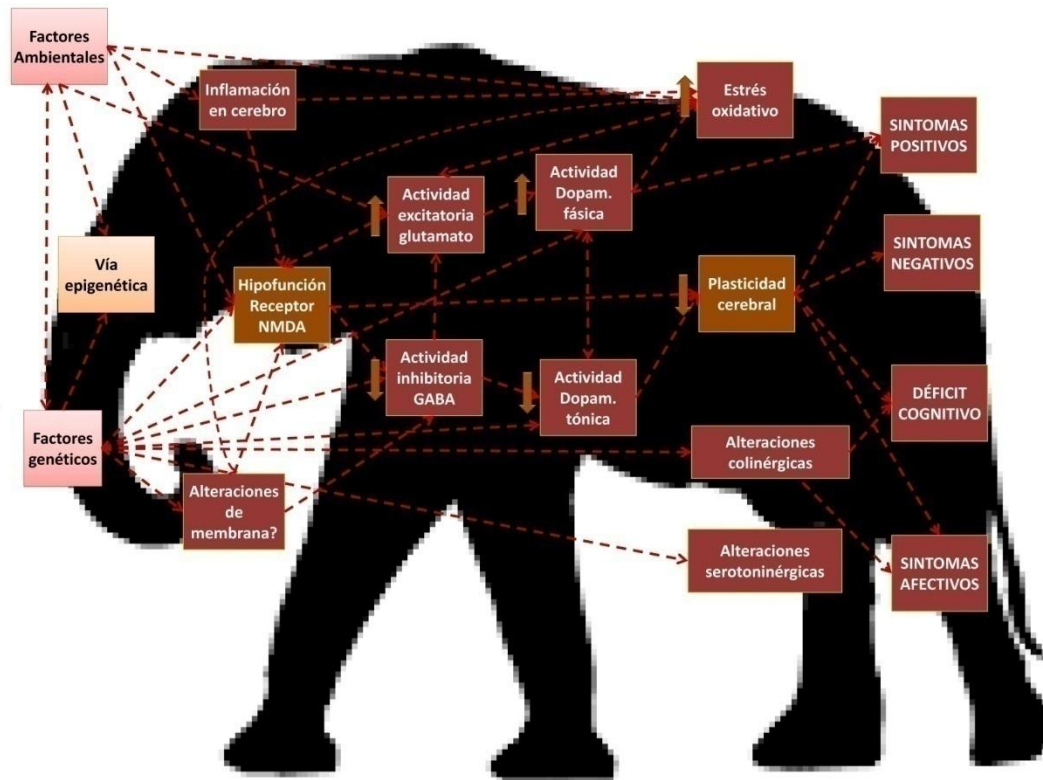
1.1.6.4. Modelo etiológico mixto.

La explicación etiológica más plausible, sin embargo, es que los diferentes modelos e hipótesis sean observaciones de la enfermedad desde distintos ángulos y perspectivas y que, al fin y al cabo, puedan integrarse en un todo común. Ante una enfermedad compleja como la esquizofrenia, donde muchos factores de predisposición variados e interrelacionados puedan cooperar para dar lugar al desarrollo de la misma, no es nada desdeñable la idea de que todas esas hipótesis, o la mayoría, sean ciertas de algún modo. Keshavan y cols. (2011) recogieron las evidencias neuroquímicas y fisiopatológicas más importantes de la esquizofrenia y su heterogéneo cuadro sintomatológico en un gráfico con forma de elefante, aludiendo a la parábola de origen hindú de *Los ciegos y el elefante*. En ella, un grupo de ciegos palpa el cuerpo del elefante (la realidad) para ver cómo es, y al comparar sus observaciones comprueban que son diferentes y que tienen una visión distinta de la realidad cuando la están observando desde distintos puntos de vista. La visión más verosímil de la esquizofrenia podría requerir relacionar las observaciones recopiladas estos años en un modelo de “realidad” semejante, y el trabajo de investigación futuro necesitar centrarse en la búsqueda de esos vínculos entre ellas (**figura 1.3**).

Bajo esta visión global, tendrían cabida varios modelos interrelacionados. Variantes genéticas y factores de predisposición ambiental se relacionan fuertemente con la hipofunción del receptor glutamatérgico NMDA, el cual, mediante la inducción de una disfunción gabaérgica, puede provocar una hiperexcitación del sistema glutamatérgico, cuya excitotoxicidad ya se ha postulado como un factor a tener en cuenta en el declive degenerativo y cognitivo de la enfermedad. En esto también influiría la interacción de los mediadores de la respuesta inflamatoria, como citoquinas o interleuquinas, con estos receptores (Müller y Schwartz, 2008). A su vez, la disfunción glutamatérgica estaría detrás de las alteraciones en las vías dopaminérgicas y de la reducción en la plasticidad neuronal (Insel, 2010). Numerosos estudios recientes apuntan a que factores ambientales podrían estar contribuyendo al fenotipo esquizofrénico a través de la regulación de la expresión génica, esto es, mediante mecanismo epigenéticos. Debido a la limitación metodológica, sin embargo, esto no ha podido confirmarse.

En global, toda esta amalgama de interacciones genéticas y ambientales, vía epigenética, que provocan alteraciones inmunológicas y neuroquímicas durante el neurodesarrollo sería, en última instancia, la consecuencia de ese “elefante” sintomatológico tan heterogéneo.

Figura 1. 3. Diagrama de la etiopatogenia de esquizofrenia, mostrando las posibles conexiones entre modelos propuestos. Figura adaptada de Keshavan y cols. (2011).



No obstante, y aunque es muy probable que este sea el escenario real, todavía queda por dilucidar si estamos ante un gran elefante, varios elefantes o una habitación con muchos animales variados. Sea como fuere, lo que parece casi irrefutable es que la genética está detrás del desarrollo de la esquizofrenia como principal factor de predisposición, tal y como demuestran los estudios de heredabilidad, y que la sitúan como una enfermedad común eminentemente genética (**apartado 1.2**). Por tanto, el estudio de la genética de la esquizofrenia es una condición necesaria para poder dibujar la ilustración más adecuada que refleje la naturaleza de la enfermedad en profundidad.

1.2. GENÉTICA DE LA ESQUIZOFRENIA

Dada la heterogeneidad fenotípica de la esquizofrenia, tal y como se ha descrito en el anterior apartado, es una cuestión de vital importancia para la investigación de la enfermedad el conocer la etiología del fenotipo y poder aplicar las medidas preventivas pertinentes. En biología, el debate *nature vs nurture* es una cuestión de gran relevancia; en el cual se contraponen

la causalidad genética a la ambiental, respectivamente. Hay que recordar que el fenotipo de un trastorno, bien sea complejo o simple, es el resultado de la interacción del genotipo y el ambiente. Es la expresión de los genes en los distintos tejidos del organismo a lo largo de su vida lo que determina el fenotipo observado, y esa expresión génica se encuentra condicionada por la acción del entorno sobre dicho organismo. De esta forma, el fenotipo observado no es una correspondencia unívoca del genotipo, sino que existe un rango fenotípico denominado norma de reacción, más o menos amplio en función de la contribución del entorno, asociado a un determinado genotipo (Stearns, 1992). En las enfermedades complejas como la esquizofrenia, donde muchos factores genéticos y ambientales determinan su aparición, el escenario alcanza una complejidad mucho mayor. Por ello, desentrañar la contribución relativa de la genética y el ambiente en el desarrollo de un trastorno como la esquizofrenia no es tarea fácil.

Sin embargo, gracias al intensivo estudio en familias y al avance de nuevos métodos al respecto, se ha demostrado que la esquizofrenia es una de las enfermedades con una mayor contribución de la genética en el fenotipo, lo cual ha justificado su estudio a lo largo de las últimas décadas y nos ha permitido gozar de una perspectiva aceptable acerca de la etiología genética de la enfermedad, algo que era impensable hace unos pocos años.

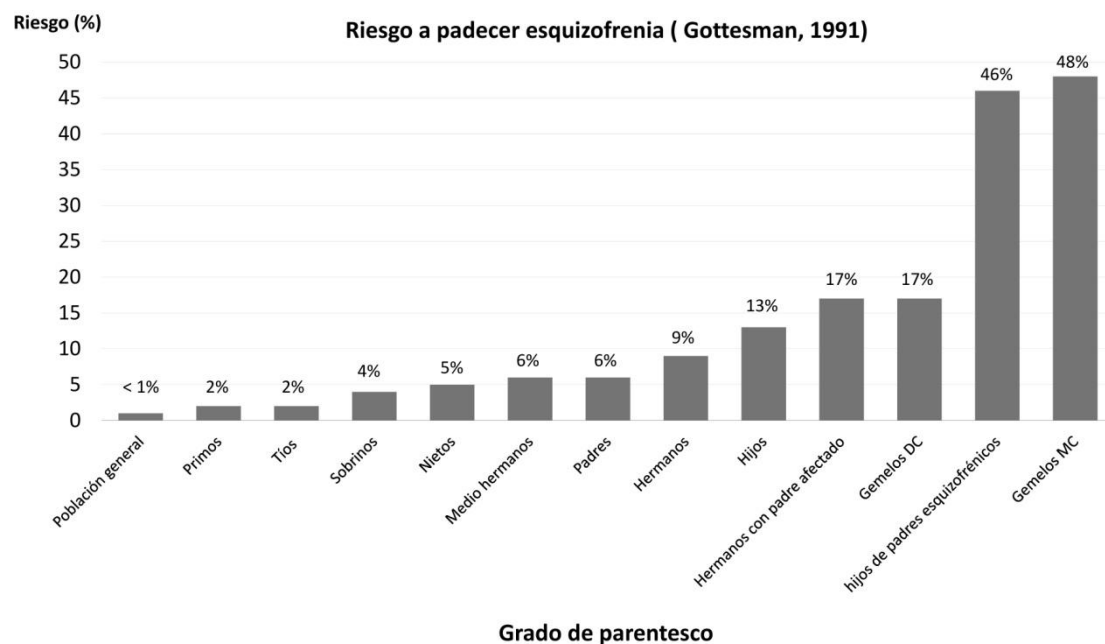
1.2.1. Epidemiología genética. Heredabilidad de la esquizofrenia

Las primeras evidencias acerca del importante papel de la genética en la esquizofrenia surgieron mediante los **estudios familiares**. Los miembros de una misma familia muestran una semejanza en su genoma que puede predecirse en función de dicho grado de parentesco. Por ejemplo, las parejas de hermanos de mismos padres comparten un 50% de su genoma, mientras que los nietos-abuelos un 25%. De este modo, se pueden organizar en diferentes grados de parentesco a los que se les espera un porcentaje determinado de compartición fenotípica en función de la compartición genética. Si un trastorno complejo tiene base genética debe mostrar un riesgo mayor en familiares de individuos afectados, y en un grado en función del parentesco, que en la población general. El riesgo a padecer esquizofrenia en los miembros familiares de un individuo afecto se encuentra claramente correlacionado con el grado de parentesco del mismo (**Figura 1.4**). Además, también ponen de manifiesto la existencia de comorbilidad con otros trastornos psiquiátricos (Riley y Kendler, 2006), de modo que es más frecuente la aparición de individuos con trastorno bipolar o autismo en familias en las que un miembro padece esquizofrenia.

La mayor evidencia del peso de la genética en familias la constituyen los **estudios de gemelos**. En este sentido, los gemelos dicigóticos (DC), que comparten el 50% de los genes, y los gemelos monocigóticos (MC), con un 100% de la variabilidad genética en común y donde la

variación fenotípica solo puede ser debida al ambiente, presentan un riesgo a padecer esquizofrenia que se eleva a un 17% y a un 46 - 50 % (Cardno y Gottesman, 2000; Allan y cols., 2009), respectivamente. En comparación, el resto de familiares de primer grado presentan un riesgo medio aproximado de 11%, con una concordancia genética del 50%, y los de segundo grado de un 4% con un 25% de la variabilidad genética compartida.

Figura 1. 4. Representación de los porcentajes de riesgo de padecer esquizofrenia frente al grado de parentesco. Figura adaptada de Gottesman (1991).



Los datos ofrecidos por Gottesman (1991) revelan, más allá de la innegable contribución genética, algunos aspectos importantes a considerar. En ellos se observa que existen claras diferencias entre el grado de parentesco y el riesgo a padecer esquizofrenia, quedando patente que no hay una distribución homogénea al respecto. Un caso muy ilustrativo es la comparación de gemelos DC y hermanos. Ambos comparten un 50% de su variabilidad genética, pero el riesgo a padecer esquizofrenia cuando su relativo padece esquizofrenia es de un 17% y 9%, respectivamente, lo cual pone de manifiesto la vital importancia del ambiente compartido durante el desarrollo y, en este caso, también el entorno prenatal (Lewis & Levitt, 2002; van Os, 2010). Es por ello que una concordancia superior en gemelos MC que en DC no demuestra la cuantía del efecto genético. El quela familia proporcione a cada generación tanto sus genes como su ambiente, sigue siendo una traba metodológica. Muchos caracteres se presentan en familias a causa del ambiente familiar compartido. Esto tiene una relevancia mayúscula cuando se trata de trastornos conductuales, como el autismo o la esquizofrenia, que

dependen sobremanera de la crianza en la infancia. No tener en cuenta el ambiente familiar compartido a la hora de demostrar la base genética de una enfermedad sería un grave error, por lo que se requiere algo más que una mera tendencia familiar para probar la base genética de los trastornos no mendelianos.

Una manera de soslayar esos inconvenientes serían los **estudios de adopción**. De esta manera, se pueden comparar los resultados de incidencia de enfermedad en parientes biológicos criados en diferentes sitios por adopción con el riesgo de recurrencia en un grupo control de familias no adoptadas. Un ejemplo es el realizado por Tienari y cols. (1994), en el que se demostraba que la agregación familiar de la esquizofrenia existía principalmente por la compartición genética y no por el ambiente compartido. El experimento ideal consistiría en gemelos MC que se separan al nacer y se crían en ambientes separados. De este modo, sí se podrían escindir al máximo las causas genéticas de las ambientales. En este sentido, los casos recogidos desde hace ya décadas también revelaron que los gemelos MC separados al nacer y criados en ambientes adoptivos diferentes siguen manteniendo el riesgo (Gottesman y Shields, 1982). Pese a las complicaciones metodológicas, los estudios de adopción en esquizofrenia han establecido y confirmado el importante papel de los factores genéticos en la etiología de este trastorno (Prescott y Gottesman, 1993; Ingraham y Kety, 2000).

Gracias a estos estudios, se ha podido llegar a cuantificar la importancia de la genética en el desarrollo de la esquizofrenia. A partir del grado de concordancia fenotípica observada y esperada en función del grado de parentesco y la corrección con los estudios de adopción se puede estimar la **heredabilidad**, que se define como el porcentaje de varianza fenotípica atribuible a la varianza genética.

La heredabilidad es, por tanto, un concepto estadístico característico de cada población en la cual se hayan estimado las varianzas ambientales y genéticas que conforman la varianza fenotípica. Los factores ambientales y genéticos que contribuyen al fenotipo pueden variar de una población a otra, lo cual impide, teóricamente, extrapolar un valor de una población a otra, aunque sea sorprendente el grado de concordancia en distintas poblaciones que existe en este parámetro cuando se estudian enfermedades complejas. Por otro lado, este valor no tiene por qué ser constante a lo largo del tiempo en tanto que los factores ambientales pueden cambiar y también, por tanto, la contribución relativa de la varianza genética a la varianza fenotípica total (Wray y Visscher, 2008).

Recientemente, también se ha propuesto la estimación de la heredabilidad mediante el análisis de la variación de la herencia genética por *identidad por descendencia* en vez de las correlaciones fenotípicas en familiares (Visscher y cols., 2006). En el caso de hermanos, la identidad genética por descendencia es de un 50%, con una desviación estándar de un 4% (lo

cual quiere decir que la identidad oscilará entre un 40 y un 60% aproximadamente). Utilizando marcadores genéticos se puede determinar si la concordancia genotípica entre hermanos se correlaciona con la identidad, caso en el cual se demostraría la relevancia de la genética en el fenotipo observado.

Desde un punto de vista matemático, la *heredabilidad en sentido amplio* (H^2) se describe como la varianza genética (V_G), dividida entre la varianza fenotípica (V_P), tal y como se muestra en la fórmula siguiente (Byers, 2008).

$$H^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_A} = \frac{(V_{Ad} + V_D + V_E)}{(V_{Ad} + V_D + V_E) + (V_{Aesp} + V_{Agen} + V_{GxA})} \cong \frac{V_{Ad}}{V_P} = h^2$$

La varianza genética la constituye el sumatorio de varianza aditiva (V_{Ad}), varianza de dominancia (V_D) y varianza por epistasia (V_E). La primera se refiere a la desviación fenotípica resultado de la presencia de unos alelos u otros en los individuos de la población. La segunda describe la posible relación de dominancia entre alelos del mismo locus, de forma que en un caso de $V_D = 0$, habría codominancia entre ambos. Finalmente, la tercera se refiere a la interacción entre alelos de distintos *loci* (epistasia), pudiéndose desglosar a su vez las interacciones debidas a dos o más *loci*. No obstante, hay numerosas evidencias de que V_{Ad} es la principal contribución a la varianza genética (Hill y cols., 2008), siendo mucho mayor que la relativa a dominancia o epistasia, por lo que una aproximación más utilizada para el cálculo de la *heredabilidad en sentido estricto* (h^2) es el que utiliza los factores aditivos como varianza genética.

La varianza ambiental (V_A) la componen la varianza específica de individuo (V_{Aesp}), que explica la variación residual por factores ambientales individuales (se correspondería con el ruido de fondo), la varianza ambiental general (V_{Agen}), que se refiere a la varianza fenotípica debida a condiciones ambientales que afectan a la población en general y es la principal contribución, y la varianza de interacción del genotipo con el fenotipo (V_{GxE}), que explica la diferente respuesta al ambiente en función del genotipo de cada individuo (Interacción gen – ambiente). En este sentido, es importante resaltar que la contribución de factores ambientales que interactúan con el material genético durante el desarrollo, sin provocar un cambio en la secuencia de pares de bases nitrogenadas mediante mecanismos epigenéticos como la acetilación o metilación de histonas y de ADN, y con la consecuente alteración de la varianza fenotípica, se integrarían dentro de la varianza ambiental (V_{Agen}).

En esquizofrenia, la heredabilidad estimada gracias a todos estos estudios comentados oscila entre un 64% (Lichtenstein y cols., 2009) y un 80% (Sullivan y cols., 2003). Tal y como se señalaba anteriormente, la heredabilidad es un parámetro poblacional, así que variaciones en estimaciones sobre poblaciones diferentes pueden variar por las varianzas ambientales y genéticas que actúen sobre ellas. Además, es necesario señalar que la estimación de la heredabilidad en humanos, dada la complejidad fenotípica de la enfermedad estudiada y la dificultad metodológica en comparación con organismos de experimentación, está sujeta a un error considerable que podría explicar estas diferencias. En cualquier caso, hablamos de una heredabilidad muy alta donde el papel de la genética es clave en el desarrollo de la enfermedad, lo que además sitúa a la esquizofrenia como una de las enfermedades complejas más genéticas que existen. Dada la prevalencia aproximada del 0,5% y la heredabilidad en torno al 64-80%, la proporción de casos esporádicos (más de un 83%) es superior a la de casos con algún familiar afectado, lo que en principio podría resultar sorprendente (Yang y cols., 2010), pero refleja la complejidad de una enfermedad compleja de naturaleza poligénica.

1.2.2. Estudios genéticos en esquizofrenia

Una vez demostrado el papel fundamental de la genética en el desarrollo de la esquizofrenia, el siguiente paso para conocer la naturaleza de la enfermedad es el de conocer qué variantes genéticas son las responsables, en mayor o menor medida, de la aparición del fenotipo esquizofrénico. Dado que un 99,9% del material genético es compartido, la respuesta debería encontrarse en el 0,1% que nos distingue genéticamente y nos hace más o menos susceptibles a padecer esquizofrenia (International HapMap Consortium, 2005).

Sin embargo, la alta heredabilidad demostrada en los numerosos estudios en familias no ha servido para explicar dos factores muy importantes. En primer lugar, no se ha demostrado la existencia de un patrón de herencia. La conclusión a la que se ha llegado es que la esquizofrenia es una enfermedad multifactorial, con muchos factores genéticos y ambientales responsables de su aparición, y compleja, donde el patrón de herencia no sigue ningún modelo conocido. Esto es también similar en el caso de otros trastornos psiquiátricos (O'Donovan y cols., 2009).

En segundo lugar, otra característica no explicada y sujeta a debate es su elevada prevalencia en la población (Messias y cols., 2007), mucho mayor que en cualquier enfermedad monogénica, y que, además, se asocia a una muy significativa reducción del éxito reproductivo (Bundy y cols., 2011). Estos dos aspectos son el origen de las hipótesis sobre la que se sustentan los estudios genéticos en esquizofrenia (Porteous, 2008).

En este sentido, son dos las principales hipótesis acerca de la naturaleza genética de la esquizofrenia que se han propuesto para tratar de explicar la elevada prevalencia y herencia compleja. Por un lado, muchos autores creen que la alta frecuencia de la enfermedad en la población se debe a que son muchas variantes genéticas comunes y de efecto menor las responsables, en su conjunto, de la aparición del trastorno (Lander, 1996). Bajo esta hipótesis, la ausencia de un patrón de herencia determinado se explicaría por el hecho de que las variantes genéticas que producen riesgo de padecer la enfermedad estarían ampliamente distribuidas por la población y la susceptibilidad sería resultado de su efecto conjunto e interacción entre ellas. Por otro lado, otros autores defienden que la alta prevalencia de la esquizofrenia en la población se debe principalmente a la aparición de variantes raras de susceptibilidad que, por su efecto deletéreo, no alcanzan una frecuencia elevada en la población pero no se llegan a eliminar completamente, segregando a baja frecuencia (Manolio, 2009).

No obstante, el actual debate que contrapone el modelo de *enfermedad común - variante común* al de *enfermedad común - variante rara* es la consecuencia de los resultados obtenidos en los diferentes estudios que pudieron llevarse a cabo a lo largo de los últimos años. Es importante resaltar que, en este sentido, el desarrollo tecnológico siempre ha sido el factor limitante para la obtención de resultados que, en última instancia, han permitido generar nuevas hipótesis a raíz del éxito o fracaso de los mismos. La aparición y mejora técnica de las distintas plataformas tecnológicas, así como la finalización de proyectos masivos como el Proyecto Genoma Humano (PGH), que dio lugar al conocimiento al completo de la secuencia del genoma humano (Lander y cols., 2001; Venter y cols., 2001), han ido permitiendo abordar el estudio de la genética de la esquizofrenia desde una óptica cada vez más ambiciosa.

1.2.2.1. Estudios de ligamiento

Los estudios de ligamiento fueron los primeros métodos utilizados para tratar de identificar variantes de susceptibilidad a esquizofrenia. Risch y cols (1990) consiguieron, mucho antes del desarrollo del PGH, identificar regiones genómicas de susceptibilidad a esquizofrenia mediante esta metodología.

La metodología consiste en el estudio de marcadores genéticos, elemento genético con una ubicación física identificable (*locus*) y cuya herencia se puede rastrear, en varias familias con varios sujetos afectos (al menos 2) mediante la caracterización genotípica y fenotípica de individuos de varias generaciones de dichas familias. La distribución de estos marcadores genéticos a lo largo de todo el genoma permite conocer si cosegregan conjuntamente con la enfermedad y la forma en que se transmiten a través de las generaciones, lo cual puede dar

pistas acerca del número de *loci* implicados y de su localización. Los genes implicados se transmitirían junto con los marcadores encontrados, en función de **ligamiento** existente entre ambos (que, a su vez, depende de la frecuencia de recombinación), de modo que un mismo *locus* podría ser una región larga que albergase varios genes.

El estadístico utilizado para evaluar la existencia de ligamiento es el **LOD score** o puntuación LOD (Z). Se trata de un estimador que nos permite detectar ligamiento al estimar la frecuencia de **recombinación** más probable entre genes en la especie humana. Se calcula como el logaritmo de la razón entre la probabilidad de la descendencia dada la existencia de ligamiento con frecuencia de recombinación Θ respecto al caso sin ligamiento.

$$Z = \log_{10} \frac{\text{Probabilidad de un genotipo a una determinada frecuencia de recombinación } \Theta}{\text{Probabilidad del mismo genotipo suponiendo segregación independiente } (\Theta=0,5)}$$

Dada su escala logarítmica es factible sumar los diferentes valores de Z obtenidos para diferentes familias (asumiendo el mismo modelo genético), y así evitar los problemas derivados de la escasa potencia estadística de considerar la descendencia de una sola familia. Lo que interesa es el valor del parámetro Θ que maximiza Z. Usualmente, se aceptan como evidencia de ligamiento valores de $Z \geq 3$, y valores de $Z < -2$, que representarían un ligamiento negativo. Hay que tener en cuenta que para la mayoría de las estadísticas se utiliza el umbral $p < 0.05$, pero se es tan estricto (1000:1 razones de probabilidades) por la improbabilidad de que dos locus, elegidos al azar, deban enlazarse y segregar ligados. La fracción de recombinación más probable sería aquella en la que la calificación *lod* es más alta.

Cuando diversos marcadores de una región candidata muestran ligamiento con la enfermedad, puede ser posible determinar la fase de ligamiento para aquel segmento cromosómico. Estas particulares combinaciones cromosómicas de alelos se conocen como **haplotipos**, y pueden verse como unidades de transmisión similar a un alelo ya que existen como tales por la ausencia de recombinación en ellas. Por tanto, llega a ser posible trazar regiones cromosómicas a través de linajes sin perder sus identidades.

Los estudios de ligamiento presentan varias ventajas a tener en cuenta. Se trata de estudios generadores de hipótesis en tanto que no es necesario un conocimiento a priori de qué regiones o genes están potencialmente involucrados, sin caer en el sesgo de estudiar solo determinadas regiones por hipótesis previas. Así, aunque se desconozcan los procesos fisiopatológicos que generan una enfermedad, se pueden usar para determinar la localización de los posibles genes que están implicados en ella. El éxito de esta técnica dependerá en gran

medida, obviamente, del número de marcadores utilizados y de que todos ellos cubran la mayor parte de los haplotipos a lo largo del genoma, algo que con el desarrollo del proyecto genoma humano y del proyecto HapMap (International HapMap Consortium, 2005), proyecto internacional creado para desarrollar un mapa de haplotipos del genoma humano, se hizo posible. Por otra parte, es un estudio muy adecuado cuando existe **heterogeneidad alélica**, esto es, la presencia de múltiples variantes dentro de una región de susceptibilidad (como podría ser un gen), sin tener porque ser la causa de esa susceptibilidad una variante en particular.

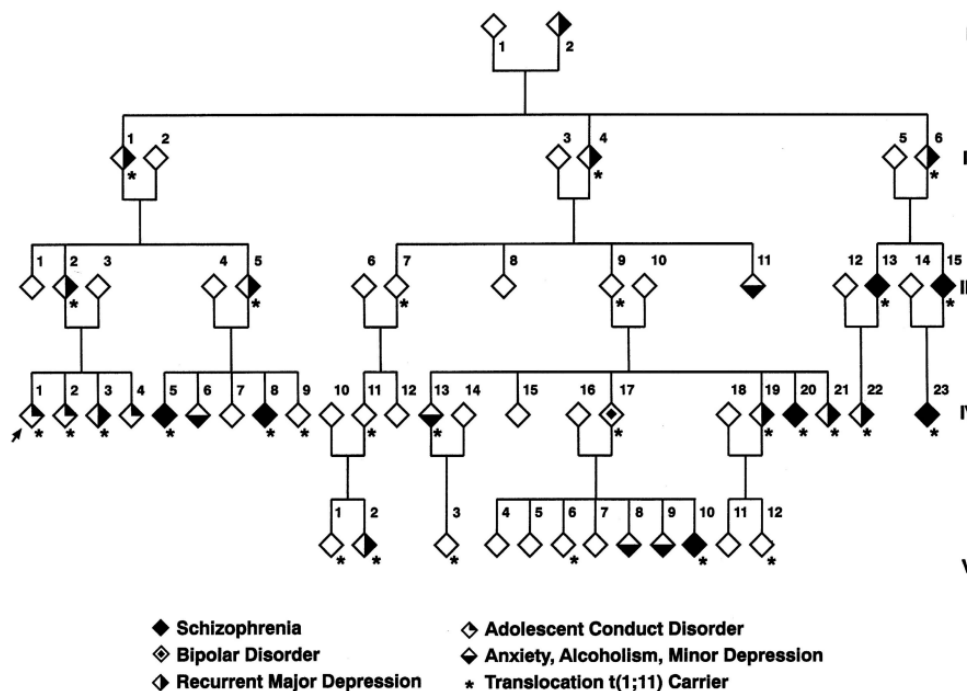
En esquizofrenia, esta estrategia condujo a la identificación de varios *loci* asociados al trastorno en familias concretas (Pulver y cols., 1994; Straub y cols., 1995; Wang y cols., 1995; Blouin y cols., 1998; Kendler y cols., 2000; Basset y cols., 2000; Gurling y cols., 2001; Stefansson y cols., 2002; MacIntyre y cols., 2003). Un estudio de ligamiento de gran relevancia en esquizofrenia es el de Millar y cols. (2000), en el que se descubrió una traslocación balanceada del gen *DISC1* que confería susceptibilidad a esquizofrenia en una familia escocesa (Figura 1.5). Además, poco más tarde quedó constatado que en la misma familia aparecían diversos síndromes psiquiátricos ligados a los portadores de la traslocación (Blackwood y cols., 2001), lo que demuestra la utilidad de estos estudios como generadores de posibles genes candidatos.

Sin embargo, pese a los numerosos resultados obtenidos, no fue posible replicarla mayoría de ellos en familias distintas a las de los estudios originales. Esta falta de reproducibilidad pone de manifiesto la **heterogeneidad de locus**, es decir, la existencia de varias regiones implicadas en el trastorno que podrían segregarse diferentemente en las distintas familias. Esta es, de hecho, una de las desventajas de esta técnica. Cuando varios *loci* están implicados, no se tiene potencia estadística suficiente para la detección de las señales de ligamiento. Otros factores limitantes asociados con la escasa potencia del método serían las familias con bajo número de miembros estudiados, la mala caracterización fenotípica y la presencia de variantes de baja penetrancia.

Un método habitualmente empleado para conseguir una mayor potencia estadística y controlar determinados sesgos es la combinación de varios estudios en uno sólo, lo que se conoce como **metaanálisis**, y en el cual el tamaño muestral es el resultado de la combinación de los originales. La ventaja principal es que se pueden detectar señales de ligamiento más débiles que en familias individuales que son comunes al conjunto de cohortes que integran el metaanálisis, pero debilitan aquellas características propias de familias particulares. Por ello, variantes como la traslocación t(1;11) en *DISC1* (Blackwood y cols., 2001) no serían significativas en un metaanálisis, ya que tal variante es específica de una familia concreta. En esquizofrenia, se han realizado varios metaanálisis (Badner y Gershon, 2002; Lewis y cols.,

2003; Ng y cols., 2009), y el más reciente, que engloba 3255 familias, detectó señales de ligamiento significativas en las regiones 1p, 2q, 3p, 5q, 6p, 8p, 10q y 16p (Ng y cols., 2009), que se detallan en la **tabla 1.2**.

Figura 1. 5. Árbol genealógico de la familia escocesa en la que se descubrió la traslocación cromosómica t(1;11) en el que se representa la incidencia de esquizofrenia y otros trastornos psiquiátricos (Blackwood y cols., 2001).



Las señales de ligamiento detectadas corresponden a *loci* que engloban varios genes y cuya señal hay que rastrear para dar lugar con el gen o genes de interés, en donde puede existir una variante o pocas de alto efecto o varias de efecto menor pero que se agrupan en el mismo haplotipo y cosegregan con la enfermedad, conformando la misma señal de ligamiento. Mediante técnicas de mapeo fino de las regiones más prometedoras se ha destacado la implicación de genes como *COMT*, *PRODH2* y *ZDHHC8* en la región 22q, *NRG1* y *PPP3CC* en 8p21-p22, *DTNBP1* en 6p24-p22, *DAAO* en 13q14-q32, *DISC1* en 1q42 o *TRAR4* en 6q21.q22 (Riley y Kendler, 2006; Owen y cols., 2005a). Muchos de estos genes también se han visto implicados en estudios de asociación o relacionados con otros genes asociados a esquizofrenia, y en muchos casos las señales de ligamiento se observan igualmente en estudios de otros trastornos psiquiátricos, sugiriendo una naturaleza genética compartida, al menos en cierto grado (Blackwood y cols., 2001). Una conclusión que se obtiene de la baja intensidad y la heterogeneidad de las señales obtenidas en estos estudios es que la esquizofrenia encaja en un

modelo es una enfermedad multifactorial compleja, con muchos *loci* de susceptibilidad y donde no se observan en ninguna familia formas de herencia mendeliana.

Tabla 1. 2. Localización y significación de los principales *loci* ligados a esquizofrenia en metaanálisis (Ng y cols., 2009). CAUC representa estudios sólo en población caucásica y ALL en todas las poblaciones.

Cromosoma	Localización (Mb)	Valor P
1p32.2-31.1	57.3-84.6	0.027 (ALL)
1p13.2-q23.3	114.6-162.1	0.008 (ALL)
2q12.1-21.2	103.3-134.0	0.0076 (ALL)
2q21.2-31.1	134.0-169.9	0.024 (ALL)
2q33.3-36.3	206.3-228.3	0.009 (ALL), 0.010 (CAUC)
3p14.1-q13.32	71.6-120.2	0.041 (ALL), 0.044 (CAUC)
5q31.3-35.1	141.8-167.7	0.0046 (ALL), 0.017 (CAUC)
5q35.1-35.3	167.7-180.4	0.028 (ALL)
6p21.31-12.1	33.9-56.6	0.044 (CAUC)
8p22-12	15.7-32.7	0.031 (ALL), 0.0006 (CAUC)
10q26.12-26.3	123.1-135.1	0.035 (ALL)
16p13.12-q12.2	13.2-51.5	0.018 (CAUC)

1.2.2.2. Estudios de asociación en genes candidatos

Los estudios de asociación pueden representar una estrategia de búsqueda más directa y potente que los análisis de ligamiento. A diferencia de los últimos, estos estudios se fundamentan básicamente en la existencia de una simple **asociación**, sin la necesidad de conocer la segregación de variantes con la enfermedad. El término asociación no es, en ningún caso, privativo de la genética, sino que hace referencia a una afirmación estadística respecto a la coocurrencia de alelos o fenotipos. El planteamiento consiste en establecer dos poblaciones (población de casos y de controles) y determinar las frecuencias alélicas y genotípicas de variantes seleccionadas para el estudio, de modo que la desviación en la distribución de frecuencias de la población de casos respecto a la de controles permita conocer el riesgo o protección que confieren esos alelos o genotipos para la aparición de la enfermedad.

Esta asociación debe ser cuantificable mediante alguna magnitud que nos informe de cuán asociado está, por ejemplo, un alelo a una enfermedad determinada, y de igual modo, si la asociación es estadísticamente significativa. Para ello, la magnitud que se suele utilizar es el *odds ratio* (OR), que es la razón de dos cocientes

$$Odds\ ratio = \frac{\frac{\text{Casos portadores del alelo A}}{\text{Casos no portadores del alelo A}}}{\frac{\text{Controles portadores del alelo A}}{\text{Controles no portadores del alelo A}}}$$

De este modo, a medida que el valor de OR se aleja de la unidad, el alelo A estaría confiriendo riesgo ($OR > 1$) o protección ($OR < 1$) a padecer la enfermedad. Este valor es, a su vez, una representación de la probabilidad de que el evento "alelo A" esté presente en el grupo de casos seleccionados o en el de controles, la cual se obtendría mediante la siguiente expresión:

$$Probabilidad = \frac{OR}{OR+1}$$

Un OR de 1 representaría una misma probabilidad de encontrar el alelo A en casos o en controles, y un OR de 2 equivaldría a una probabilidad del 66,7% de que dado un alelo A, éste se halle en un caso. No obstante, el valor de OR no nos da una idea de la significación estadística de tal asociación, ya que factores como los tamaños de muestras de enfermos y sanos influyen en ella. Por eso, Para determinar si la frecuencia del alelo analizado es significativamente diferente entre ambos grupos se emplea una prueba estadística, habitualmente la *prueba Chi-cuadrado* (χ^2) o la *regresión logística*. En estas pruebas, la hipótesis nula es la igualdad de frecuencias y que las diferencias encontradas son debidas al error aleatorio inherente a la distribución. La hipótesis alternativa sería la existencia de asociación ya que las diferencias encontradas no se pueden explicar por azar para un nivel de significación determinado (α). El valor de α habitualmente empleado es 0,05, lo cual quiere decir que si un contraste de hipótesis por una prueba como la χ^2 proporciona un valor p inferior a 0.05, la hipótesis nula es rechazada, siendo el resultado estadísticamente significativo.

En el caso de la esquizofrenia, donde la ausencia de resultados muy significativos en los estudios de ligamiento, en comparación con enfermedades mendelianas, demostraron la existencia de una clara **heterogeneidad de locus** con múltiples genes involucrados, los estudios de asociación representaron un lógico paso a seguir para tratar de determinar la naturaleza genética de la enfermedad de forma más precisa. Dada la certeza de la existencia de variantes de predisposición de efecto más bajo, los estudios de asociación representaron a priori una estrategia mucho más potente que los estudios de ligamiento (Risch y Merikangas, 1996). Todo esto fue posible gracias a grandes inversiones internacionales que financiaron proyectos masivos como el Proyecto Genoma Humano, dando lugar al conocimiento de muchos polimorfismos a lo largo del genoma humano y posibilitando testar la asociación en variantes presentes en muchas regiones y/o genes susceptibles de estar relacionados con la enfermedad. Además, en comparación con los estudios anteriores, los de asociación conllevan una recogida de muestras mucho más sencilla, donde lo importante es la caracterización fenotípica y

poblacional de los casos y los controles y, en ningún caso, la búsqueda de familias concretas con varios individuos enfermos a lo largo de varias generaciones.

No obstante, por contrapartida, los estudios de asociación, que se fundamentan en la existencia de variantes alélicas de susceptibilidad, son ineficaces frente a la existencia de distintos alelos de un mismo gen asociados con la enfermedad (**heterogeneidad alélica**), algo que no ocurre en el caso de los estudios de ligamiento.

Se debe de tener claro que la asociación, como un concepto poblacional que va más allá de la propia genética, puede ocurrir en la población por determinadas causas:

- La primera, y más deductiva, es la **causa directa**. Esta implica tener un alelo A torna a una persona más susceptible a padecer la enfermedad D, aunque, evidentemente, este hecho no sea ni necesario ni suficiente para que se desarrolle la enfermedad.
- En segundo lugar, dado que la población está contenida por varios subgrupos poblacionales distintos, tanto la enfermedad como el alelo A podrían ser más o menos frecuentes en un determinado tipo. Es lo que se conoce como **estratificación poblacional**, y es una de las principales fuentes de error y por ello, uno de los factores a tener en cuenta cuando se diseña un estudio de asociación. Constituye el ejemplo más claro de que la asociación es una característica de la población y que, aunque es una gran estrategia para desvelar causalidad genética, es imprescindible trabajar con muestras de casos y controles del mismo estrato poblacional sin incurrir en ningún sesgo en el proceso de selección.
- Por último, la asociación podría ser **indirecta**, al estar **causada por DL**. Esto es que otra variante diferente a la estudiada fuese la responsable de la asociación con la enfermedad y además se encontrase ligada a la primera. De este modo, las dos variantes serían marcadores cercanos que coseguran juntos en la población, dando la misma información, y siendo indistinguibles desde el punto de vista causal. Hay que tener en cuenta que el DL sí es un concepto genético, en tanto que la asociación no era sino uno estadístico. Por ello, deben ser separados conceptualmente, y aunque están relacionados, el ligamiento por sí mismo no produce una asociación en la población general sino que crea asociaciones dentro de una familia, pero no entre personas no relacionadas. Un ejemplo podría ser que un marcador X estuviera ligado a un determinado gen responsable de una enfermedad, pero sin embargo no estar asociado a la misma por la sencilla razón de que su frecuencia en la población general es similar a la de la población afectada.

Frente a las ventajas antes mencionadas, los estudios de asociación presentan algunos inconvenientes. Por un lado, son una herramienta muy útil para desvelar la asociación de

variantes con una frecuencia elevada en la población (polimorfismo), pero conseguir desvelar la asociación de variantes raras requeriría de tamaños muestrales de varios miles de individuos para alcanzar suficiente potencia estadística, algo que es muy difícil de conseguir y económicamente prohibitivo en la mayoría de los casos. Esto contrasta con lo indicado en los estudios de ligamiento para estudiar una variante rara de reciente aparición y que segrega en una o pocas familias, demostrando que ambas estrategias nunca debieran verse como la evolución de una respecto a la otra, sino como métodos compatibles y complementarios incluso actualmente. Por otra parte, los estudios de asociación requieren una hipótesis previa para la selección de un gen o genes candidatos sobre los que se va a estudiar su asociación con la esquizofrenia, algo que tampoco ocurría en los estudios de ligamiento. Debido a esto, el caso más sencillo consiste en trabajar con genes candidatos por los conocimientos previos acerca de aspectos moleculares del gen como su expresión genética en el cerebro, su función biológica, la ruta metabólica en la que está implicado o dianas de acción de psicofármacos. En este caso hablaríamos de **genes candidatos funcionales**. Fundamentar esta hipótesis previa es un hecho crucial, ya que permite disminuir posibles falsas asociaciones por azar, y reducir el número de pruebas estadísticas realizadas. No obstante, la selección no es tarea fácil, entre otras cosas, porque un 86% de todos los genes se expresan en algún momento a lo largo de la vida humana en el cerebro (Kang y cols., 2011).

La otra alternativa consistiría en realizar un estudio en aquellas zonas donde se ha detectado **ligamiento previo** en familias. Hay que tener en cuenta que estas regiones pueden contener decenas de genes por lo que, además, podrían servir como complemento para confirmar esa implicación y afinar la posición. Es por esta razón por la que el marcador genético idóneo en estos estudios es el **polimorfismo de nucleótido simple (SNP)**, ya que son mucho más abundantes que los polimorfismos clásicos (RFLPs, microsatélites, minisatélites,...), y por ello se pueden barrer más pormenorizadamente regiones del genoma en las que se encuentran señales de ligamiento. Así, se calcula que hay al menos 10 millones de SNP en el genoma humano, lo que representa 1 SNP/300 pares de bases. Además, la tecnología actual permite analizar múltiples SNP en gran número de muestras de forma simple y precisa (Syvänen, 2001). Los SNP son variantes que han alcanzado una frecuencia elevada en la población ($> 1\%$), por lo que se consideran comunes, e implican el cambio en un nucleótido de una base nitrogenada por otra o la delección o inserción del mismo. Son la principal fuente del 0,1% de variabilidad genética en el genoma humano y se encuentran distribuidos a lo largo del mismo pero de forma no homogénea, siendo diferente su frecuencia en determinados genes, en las regiones exónicas o intrónicas del mismo gen o en las regiones intergénicas. Para los estudios de asociación, los SNP son genotipados en las poblaciones de casos y controles estudiadas para probar su

implicación en la enfermedad mediante la diferencia significativa de la frecuencia en la que se distribuyen sus alelos.

El cambio nucleotídico que constituye un SNP puede tener mayor o menor repercusión funcional, lo que determina la estrategia a utilizar para la búsqueda de asociación. En primer lugar, la estrategia más intuitiva para un estudio de asociación sería la del **método directo** (Risch, 2000). En este caso, de lo que se trata es de estudiar únicamente SNP u otros polimorfismos sobre los que existan evidencias de funcionalidad y que, por tanto, esa variación en la secuencia puedan tener repercusiones que pudieran estar implicadas en la etiología de la enfermedad. Con este propósito, lo más común es analizar polimorfismos que alteren la secuencia de la proteína que expresa el gen estudiado (**SNP no sinónimos**), dando lugar a un cambio de un aminoácido (*SNP missense*) por otro o a la terminación prematura de la traducción por la aparición de un codón de parada (*SNP nonsense*). Sería menos frecuente, según esta estrategia, el caso de utilizar variantes aparentemente neutras, como SNP codificantes que no alteran la secuencia proteica (**SNP sinónimos**) o SNP intrónicos. Sin embargo, en los últimos años han ido cobrando más importancia las variantes en regiones no codificantes debido a la demostración del relevante papel regulador de lo que anteriormente era conocido como "ADN basura", por lo que las variantes que afectan en estas regiones a lugares de unión de factores de transcripción en promotores, zonas de alta conservación o regiones con modificaciones epigenéticas, pueden ser igualmente funcionales (Weinhold y cols., 2014; Roussos y cols., 2014). Actualmente, además, existen numerosas herramientas para la predicción *in silico* de la funcionalidad de una variante, como es el caso de PolyPhen (Adzhubei y cols., 2010) o SIFT (Ng y Henikoff, 2003) en regiones codificantes o RegulomeDB (Boyle y cols., 2012) y HaploReg (Ward y Kellis, 2012) en regiones reguladoras, entre otras. No obstante, la falta de evidencias funcionales acerca de muchas de las variantes estudiadas sigue siendo el factor limitante de esta estrategia.

Una alternativa que permite soslayar estos inconvenientes es el **método indirecto**, que se apoya en el DL para analizar varias variantes que cubran un gen o región de interés, sin tener en cuenta la funcionalidad de las mismas. Esto requiere que los polimorfismos que se analizan cubran la variación genética común del *locus* de interés.

La forma en que se debe de entender el DL sobre el que se apoyan los estudios de asociación es que, en las poblaciones que derivan de un ancestro común en el que ocurre una mutación, existirán bloques de secuencia genética más o menos grandes que se hereden de forma conjunta de una generación a otra (**haplotipos**), y así sucesivamente, sin que se vean afectados por la recombinación durante la meiosis. Esta es la razón por la que existe ligamiento entre algunas variantes en poblaciones que derivan de un ancestro común, y que alcancen a

compartir esa mutación fundadora. La existencia del DL origina una estructura del genoma humano caracterizada por la presencia de bloques haplotípicos (Patil y cols., 2001; Gabriel y cols., 2002; Costas, 2005), regiones del genoma que presentan una baja diversidad haplotípica (pocos haplotipos a alta frecuencia) y un alto DL entre las distintas variantes del bloque. Esta estructura de bloques haplotípicos varía entre distintas poblaciones debido a sus historias demográficas específicas (cuellos de botella poblacionales, expansiones poblacionales, endogamia, mezcla con otras poblaciones, etc) Así, las poblaciones del África subsahariana presentan la mayor diversidad haplotípica debido a que la salida del hombre moderno del continente africano para colonizar los demás territorios estuvo asociada a un cuello de botella (reducción del tamaño poblacional) que provocó una pérdida de haplotipos por azar. Como consecuencia, los bloques haplotípicos son de mayor longitud en poblaciones no africanas (Weiss y Clark, 2002; Tishkoff y Verrelli, 2003). Encontrar DL en poblaciones africanas es más complicado dado que esas regiones ya han sufrido mucha más recombinación. Como es evidente, las poblaciones ideales para realizar estudios de DL serían poblaciones aisladas originadas a partir de un pequeño grupo de fundadores y lo más tarde en el tiempo como fuera posible, de modo que no hubiera suficientes generaciones como para perder esos haplotipos comunes en la población. Islandia, por ejemplo, es un buen modelo de población a tales efectos.

El DL no es, por tanto, una propiedad cualitativa sino un parámetro cuantificable. Existen varias formas de medir el DL (Devlin y Risch, 1995). La primera, y más sencilla, es **D**. Este coeficiente mide diferencias de frecuencias de haplotipos observadas y esperadas a partir de las de cada alelo. Por ejemplo para dos SNP X e Y donde A y B son los alelos de mayor y menor frecuencia, respectivamente, D se calcularía mediante la siguiente expresión.

$$D = f(X_A \cap Y_A) - f(X_A)f(Y_A)$$

D tiene el inconveniente de que depende de las frecuencias de los distintos alelos, por lo que no es útil para comparar el DL entre distintas regiones. Por ello, se han diseñado nuevas medidas de DL entre dos marcadores, siendo las más usadas **D'** y **r²**. D' se calcula dividiendo D entre el valor máximo que pueda alcanzar, de modo que su valor oscila entre 0 (no hay DL) y +1 o -1 (máximo desequilibrio de ligamiento).

$$D' = f(X_A \cap Y_A) - f(X_A)f(Y_A) / D_{\max}$$

Los valores positivos o negativos dependen de si el alelo mayoritario de X está asociado al mayoritario o minoritario de Y, respectivamente. Valores de $0,33 < D' < -0,33$ se consideran asociaciones aparentes. Por otro lado, **r²** es una medida más restrictiva que D', ya que solo alcanza un valor máximo cuando cada uno de los alelos de un SNP está asociado a un único alelo del otro, mientras que D' no tenía en cuenta la frecuencia de ambos alelos. Además, es de

más fácil interpretación, ya que para una potencia y umbral de significación determinados, el aumento en tamaño de muestra necesario para un DL reducido es inversamente proporcional a r^2 . De este modo, si se necesitan 1000 casos para detectar asociación en condiciones ideales ($r^2 = 1$), se necesitarían 5000 para la situación probable de $r^2 = 0,2$ (Zondervan y Cardon, 2004).

$$r^2 = D^2 / (f(X_A)f(Y_A)f(X_B)f(Y_B))$$

Gracias a la estructura genómica en bloques haplotípicos, es posible genotipar solo un determinado número de SNP a lo largo de la región o gen de interés de forma que solo este grupo de SNP, denominados **tagSNP**, sean informativos, en la mayor medida posible, de todos los haplotipos presentes. De esta forma, si una variante causal se halla asociada a un tagSNP que representa su haplotipo, la asociación sería igualmente detectada. El problema evidente que presenta esta estrategia indirecta es que requiere el conocimiento de la estructura haplotípica del gen o genes candidatos. Es por ello que en el año 2002 comenzó el proyecto HapMap con el propósito de caracterizar el DL existente a lo largo del genoma (International HapMap Consortium, 2005) para que se pudiera disponer de una base de datos pública con tal información para futuros estudios de asociación. Esto, indudablemente, supuso un enorme avance para los estudios de asociación (International HapMap 3 Consortium, 2010), y algo vital para los estudios de asociación de genoma completo (GWAS) que se analizan en el siguiente apartado.

Los resultados de asociación en genes candidatos representaron un gran avance para el estudio de muchas enfermedades pero, sin embargo, en el caso de la esquizofrenia, sus resultados no fueron todo lo fiables que se suponía, llegando en numerosas ocasiones a hallazgos contradictorios, como en los casos de *COMT*, el gen de la enzima *catecol o-metil transferasa* (Strous y cols., 1997; Glatt y cols., 2003) o en *DTNBP1*, el gen de la *disbindina* (Vilella y cols., 2008; Mutsuddi y cols., 2006), ambos genes ya implicados en señales de ligamiento en estudios previos. En general, la mayor parte de los esfuerzos en esquizofrenia estuvieron centrados en la búsqueda, relativamente infructuosa, de asociación con genes implicados en los sistemas de neurotransmisores implicados en la acción de antipsicóticos de primera y segunda generación. Pese al alto número de estudios en genes candidatos, ningún gen mostró una asociación incontestable como sí había ocurrido en otras enfermedades complejas. Casos importantes como los del gen *APOE* en Alzheimer, cuya asociación inicial supuso la justificación del modelo de enfermedad común – variante común (Saunders y cols., 1993) o la asociación del gen *MTHFR* con niveles de homocisteína (Frosst y cols., 1995), no tuvieron lugar en los primeros estudios realizados en esquizofrenia. Esto llevó a la utilización de metaanálisis para tratar de ganar mayor potencia estadística y desvelar asociaciones de variantes de riesgo menor. Las más de 1000 asociaciones reportadas en esquizofrenia se agruparon en 118

metaanálisis, que se recogen en la base de datos de *SZgene database* (Allen y cols., 2008). En ella se pueden ver las 24 variantes que mostraron significación estadística a lo largo de los genes *APOE*, *COMT*, *DAO*, *DRD1*, *DRD2*, *DRD4*, *DTNBP1*, *GABRB2*, *GRIN2B*, *HP*, *IL1B*, *MTHFR*, *PLXNA2*, *SLC6A4*, *TP53* y *TPH1*. El valor de OR medio de las variantes aquí recogidas es de 1,23, un valor muy bajo teniendo en cuenta la selección tan profusa llevada a cabo. Sin embargo, y basándose en estrictos criterios epidemiológicos (Ioannidis y cols., 2008), cuatro de esos genes presentan asociación con un alto grado de credibilidad: *DRD1*, *DTNBP1*, *MTHFR* y *TPH1*.

Los escasos resultados relativos a la presencia de factores de riesgo común podrían estar causados por la mala selección de genes candidatos basada en hipótesis no acertadas o por una arquitectura genética distinta a la consistente en variación común a lo largo de estos genes analizados. Es también probable que muchas de estas irregularidades en los resultados se deban al mal diseño de estos estudios y a los errores cometidos, tanto positivos como negativos (Cardon y Bell, 2001). Estos errores se denominan **errores de tipo I**, esto es, rechazar la hipótesis nula de no asociación cuando es cierta (**falsos positivos**) o **errores de tipo II**, aceptar la hipótesis nula de no asociación cuando es falsa tanto en estudios originales como en las réplicas (**falsos negativos**), respectivamente.

Falsos positivos o errores de tipo I

El hecho de que una parte importante de los estudios de asociación no logre replicarse puede ser debido a que estas posibles asociaciones iniciales sean falsos positivos. La existencia de éstos puede tener lugar, principalmente, por tres razones.

- En primer lugar, por el **sesgo** existente hacia la publicación de estudios de asociación con resultados positivos frente a negativos, con independencia de la calidad de los trabajos (Ioannidis y cols., 2008)
- En segundo lugar, el **azar**. Se suele trabajar a un nivel de significación del 5%. Pero realmente esto indica que la probabilidad de que las diferencias observadas sean debidas al azar es del 5 %. Por tanto, si se realizan múltiples comparaciones este nivel de significación generaría un número inaceptable de errores de tipo I (se esperaría como promedio un falso positivo cada 20 pruebas). Por esta razón, actualmente el enorme número de pruebas realizadas necesita la utilización de métodos de **corrección por comparaciones múltiples**. Uno muy conocido y empleado es el *método de Bonferroni* (Dunn, 1961), en que el nuevo nivel de significación se calcula dividiendo el valor α entre el número de pruebas. No obstante, este test es altamente conservador, ya que

supone la completa independencia de los test realizados. Tras la aparición de los métodos de *False discovery rate (FDR)*, se concluyó que esta aproximación a la hora de corregir por múltiples pruebas era más adecuada y potente, siendo un referente en genética el *método de Benjamini-Hochberg* (Benjamini y Hochberg, 1995; Benjamini y cols., 2001)

- La otra causa principal que puede generar falsos positivos es la existencia de **estratificación poblacional**, tal y como se describió anteriormente. Esto es, cuando la población bajo estudio, supuestamente homogénea, es realmente una población híbrida configurada por la mezcla de individuos de distinta procedencia.

Falsos negativos o errores de tipo II

En el caso de los errores de tipo II, las dos causas principales son la poca potencia de la prueba estadística para replicar un efecto genético concreto y la incorrecta clasificación de los fenotipos.

- Gran parte de los errores de tipo II son debidos a que el tamaño muestral es insuficiente, por lo que es precisa una determinación *a priori* de la **potencia estadística** asociada al diseño que nos permita saber si es probable replicar una supuesta asociación (Ioannidis y cols., 2008). Por otro lado,
- La otra gran fuente de falsos negativos es la **clasificación incorrecta del fenotipo**, generalmente debida a la presencia de fenocopias (fenotipos de la enfermedad bajo estudio sin base genética). La principal consecuencia de la clasificación incorrecta del fenotipo es que se reduce la potencia de la prueba estadística para detectar la asociación. Para evitar este problema, debe recurrirse a un criterio diagnóstico estandarizado y a una descripción precisa de los grupos de casos y controles. Por ello, en una gran cantidad de trastornos complejos o caracteres de difícil fenotipo y, por tanto, diagnóstico, se ha propuesto el establecimiento de determinados endofenotipos (Gottesman y Goud, 2003). Estos pueden ser rasgos de personalidad, características psicológicas, cognitivas o aspectos neuroquímicos. En el caso de esquizofrenia todavía cobran mayor relevancia ya que, tal y como se describe en el primer bloque de la introducción, hay una ausencia de marcadores biológicos inequívocos de la enfermedad. El estudio de endofenotipos se ha propuesto como una buena estrategia alternativa para superar las dificultades metodológicas derivadas de la nosología en psiquiatría.

La conclusión que se puede sacar de los estudios de genes candidatos en esquizofrenia es que, pese a que sería muy injusto decir que no han aportado prácticamente nada, ya que se han consolidado implicaciones previas de algunos genes como *DRD1*, *DTNBPI*, *MTHFR* y *TPH1*, la realidad es que solo han conseguido explicar una pequeña parte de la heredabilidad de la enfermedad, por lo que en general no han resultado exitosos. Prueba de ello es que la base de datos de *SZgene* se ha dejado de actualizar desde el año 2011, muy probablemente porque los estudios de asociación de genoma completo (GWAS) representan ya una mejor y mayor perspectiva de la variación común en esquizofrenia, sin las limitaciones características de los estudios en genes candidatos.

1.2.2.3. Estudios de asociación de genoma completo (GWAS)

Los estudios de asociación de genoma completo (GWAS) representan una alternativa más potente que los de genes candidatos para el conocer la varianza genética asociada a enfermedades complejas. La aparición de esta nueva estrategia tuvo que ver con la secuenciación de la totalidad del genoma humano (Lander y cols., 2001; Venter y cols., 2001), el avance en el desarrollo de tecnologías que han permitido la genotipación de miles de variantes a un precio cada vez más asequible o la caracterización de cientos de miles de SNP (Sachidanandam y cols., 2001; Sherry y cols., 2001), que han permitido, entre otras cosas, el desarrollo del proyecto HapMap (International HapMap Consortium, 2005) y la consecuente disposición del conocimiento de la estructura haplotípica del genoma humano y de tagSNP que representasen en mayor medida estos haplotipos, indispensables para la realización de GWAS (Hirschhorn y Daly, 2005).

Los GWAS (del inglés *Genome Wide Association Studies*), se basan en el genotipado de miles de SNP que representan la mayor parte de la variación común existente. Por tanto, se sigue la estrategia indirecta por desequilibrio de ligamiento, ya descrita anteriormente, pero en la totalidad del genoma. Mediante la utilización de paneles se procede al genotipado de miles de tagSNP a lo largo de la secuencia del genoma humano. No existe una hipótesis previa que conduzca al estudio de un gen o genes en cuestión, sino que el propio GWAS será el elemento generador de hipótesis que revelará las asociaciones pertinentes sin el sesgo de comprobación de una teoría previa, como sucedía en los estudios en genes candidatos.

Estos estudios representan, por tanto, una manera más exhaustiva de analizar la genética de la enfermedad compleja sobre la base de la teoría de la **variante común - enfermedad común**, donde las variantes causales, si no genotipadas, debieran estar representadas por tagSNP de haplotipos determinados. Sin embargo, esta aplicación del concepto de asociación

indirecta por DL en la totalidad del genoma, pese al desarrollo metodológico y del conocimiento de miles de SNP, conllevó importantes limitaciones económicas, por la necesidad de genotipar miles de variantes por individuo, y metodológicas, por la necesidad de lidiar con un ingente número de test estadísticos que podrían propiciar la aparición de un alto número de falsos positivos. Estos inconvenientes, que en un principio llevaron a muchos autores a cuestionar el esfuerzo que implicaba la utilización de estas técnicas y la alta probabilidad de no obtener resultados (Weiss y Terwilliger, 2000; Blangero, 2004), fueron condicionando las estrategias a seguir durante la utilización progresiva de estos estudios a lo largo de los años.

El número de tagSNP seleccionados es un paso clave, y depende del grado de DL deseado y de la eficiencia a la hora de seleccionar SNP presentes en regiones con bajo DL (Goldstein y cols., 2003). Varios métodos estadísticos, en función de estos parámetros, fueron optimizados para tal propósito y la correcta selección de tagSNP (Johnson y cols., 2001; Carlson y cols., 2004). Una alternativa propuesta para evitar los altos costes fue la realización de GWAS mediante el análisis únicamente de **SNP missense** (Botstein y Risch, 2003), aludiendo que la variación con efectos funcionales más importante quedaría cubierta. De esta forma, ya que un gen contiene un promedio de 1-2 *SNP missense*, se podría abarcar todo el genoma tan solo mediante el análisis de entre 30000 y 60000 variantes. Sin embargo, este método presenta grandes limitaciones. Por un lado, otras variantes codificantes no estarían presentes como los *SNP nonsense* o las variantes de *splicing*. Por otro lado, es muy probable que la mayor parte de las variantes *missense* tengan un efecto funcional tan elevado que provoque que la selección negativa actúe sobre ellas, eliminándolas o reduciendo su frecuencia por debajo del umbral del 1%. Por último, dado que los estudios de asociación en genes candidatos condujeron a resultados donde las variantes presentaban efectos bajos ($OR < 1,3$), es muy probable que gran parte de la variación común se encuentre en regiones no codificantes cuyos efectos sean menores, como por ejemplo causando una pequeña reducción en la expresión de un determinado gen, y que su presencia a una frecuencia elevada no se tenga en cuenta mediante esta estrategia. No obstante, si no se tiene acceso a grandes muestras ni una capacidad económica suficiente, sigue siendo una estrategia viable que, además, en el caso de esquizofrenia, ha dado lugar a resultados positivos, como en la asociación del gen *SLC39A8* mediante GWAS de SNP no sinónimos (Carrera y cols., 2012)

Al genotipar tantas variantes, el nivel de significación estadístico necesario para afirmar que existe asociación en un estudio de GWAS debería corregirse por el número de test realizados, pues un valor de probabilidad habitual de $\alpha = 0,05$ llevaría, en caso de realizar 500000 test, a que alrededor de 25000 asociaciones apareciesen sencillamente por azar, dentro de las cuales estarían presentes las asociaciones reales. Se ha propuesto un valor de $p < 5 \times 10^{-8}$ para considerar que existe asociación (Risch y Merikangas, 1996), lo que deriva de un valor P

de 0,05 corregido mediante el método de Bonferroni para 1×10^6 test realizados. Esta forma de corregir el valor por múltiples test es muy conservadora, ya que asume independencia entre test pese a no ser demostrable, pero proporciona un alto valor de seguridad en las asociaciones encontradas.

En el caso de los GWAS, es imprescindible tener en cuenta la potencia estadística. Sin hacer una selección del tipo de variante, el requisito principal es, por tanto, el tamaño de la muestra de casos y controles. En esquizofrenia, donde una variante podría tener un valor de OR de 1,25, si su frecuencia en la población es del 15% se necesitaría unas muestras de 6000 casos y 6000 controles para alcanzar una potencia del 80% de detectar asociaciones a nivel de $p < 5 \times 10^{-8}$, es decir, un 80% de probabilidad de que se rechace la hipótesis nula cuando la hipótesis alternativa es cierta, o dicho de otro modo, de no cometer un error de tipo II. Para sobreponerse a esta limitación, existen tres principales estrategias:

- Una forma sería realizar el estudio mediante un **procedimiento de múltiples etapas**. Este consistiría en rebajar el umbral de significación al valor nominal de 0,05 en la primera fase o etapa, de modo que las muchas asociaciones que aparecen por azar (y entre las cuales, en caso de existir, estarían las verdaderas), se llevarían a una segunda e incluso tercera etapa donde en un tamaño de muestra igual o superior se genotipan exclusivamente las que hubieran pasado esta primera fase, con la consecuente corrección solo por el número de test de la segunda o tercera fase. De esta forma, se consigue no eliminar tantas posibles variantes reales en un primer análisis por Bonferroni, pero se necesita disponer de un tamaño muestral grande para poder dividir los análisis por etapas.
- Una segunda manera sería utilizar otro método de corrección por múltiples test. Dado que el método de Bonferroni no tiene en cuenta el solapamiento de haplotipos y la correlación entre marcadores, se han propuesto otros métodos menos conservadores mediante aproximaciones frecuentistas como el FDR o *False Discovery Rate* (Benjamini y Hochberg, 1995; Benjamini y cols., 2001) o bayesianas (Wacholder y cols., 2004). Otra forma muy utilizada es la **corrección por permutaciones** del valor. De esta forma, las muestras de casos y controles se mezclan al azar miles de veces, y la distribución del p valor de cada variante tras las múltiples permutaciones se compara con el obtenido inicialmente con las muestras de casos y controles separadas.
- En tercer lugar, otra estrategia sería la de la utilización de **poblaciones fundadoras**. Esta se fundamenta en el hecho de que, al usar poblaciones que han surgido recientemente por efecto fundador (es decir, que se han originado a partir de pocos individuos hace 100 o menos generaciones), los haplotipos son de mayor tamaño por la menor recombinación ocurrida en ellos desde su fundación y, por tanto, se puede cubrir

la variación común del genoma con menos SNP. Sin embargo, los primeros estudios que siguieron esta estrategia demostraron que no había diferencias notables en la estructura de DL que justificase la selección prioritaria poblaciones fundadoras (Kruglyak, 1999; Shifman y cols., 2003).

- Finalmente, una alternativa metodológica es **agrupación de las muestras en agrupaciones de muestras o pools** (Sham y cols., 2002). Esta forma de proceder presenta como principal ventaja la económica, pudiendo reducirse decenas de veces el coste del genotipado (Barratt y cols., 2002). Por otra parte, sin embargo, necesita una mayor precisión tecnológica para inferir la presencia de un alelo dentro de un *pool* de varias muestras. Esta estrategia representa una alternativa real muy válida en los casos en los que no se permisible un desembolso económico grande. Sin embargo, y dada la progresiva reducción de los costes de genotipado durante la última década, la agrupación por *pools* puede representar una alternativa más ventajosa en el caso de la secuenciación.

En esquizofrenia, los primeros estudios de GWAS resultaron muy poco satisfactorios (Kirov y col., 2009; Lencz y cols., 2007; Need y cols., 2009; O'Donovan y cols., 2008; Shifman y cols., 2008; Sullivan y cols., 2008), y solo en el trabajo de O'Donovan y cols. (2008) el gen *ZNF804A* resultó asociado a nivel de significación genómica cuando se juntaron muestras de esquizofrenia y de trastorno bipolar ($P = 9,96 \times 10^{-9}$), lo que a su vez significó una nueva demostración de la semejanza sintomatológica de ambos trastornos también debía de ir acompañada de una arquitectura genética similar.

Fue en el año 2009 cuando, tras aumentar los tamaños de las muestras de forma significativa, se publicaron tres GWAS que empezaron a revelar más resultados que los primeros intentos frustrados (Purcell y cols., 2009; Shi y cols., 2009; Stefansson y cols., 2009), demostrando que gran parte del problema en los primeros casos era la escasa potencia estadística para encontrar variantes comunes de efecto menor. Los SNP asociados implicaban a los genes *TCF4* y *NRGN* y a la región del MHC. Además, en el trabajo de Purcell y cols (2009) se confirmó la existencia de un componente poligénico (Gottesman y Shields, 1967) de 0,03 (3%) pero que, mediante la aplicación de modelos matemáticos que tuvieran en cuenta las limitaciones del estudio, se observó que podría explicar hasta un tercio de la varianza de la enfermedad (aunque en todo caso fueron estimaciones matemáticas). Este componente poligénico fue significativo también en muestras independientes de trastorno bipolar, donde se veía que explicaba un porcentaje de varianza algo menor, mientras que no existía en otras enfermedades no psiquiátricas usadas a modo de control negativo. Esto demostraba que cientos e incluso miles de variantes, sin tan siquiera asociación nominal, podrían estar actuando

conjuntamente como un factor de susceptibilidad por el cual se podría explicar gran parte de la variación común no identificada en genes específicos.

En los siguientes años, gracias a las colaboraciones y a las iniciativas para crear consorcios como el *Irish Schizophrenia Genomics Consortium* (ISGC) o el *Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium* (ahora conocido como el *Psychiatric Genomics Consortium (PGC)*), hasta 30 *loci* presentaron asociación con significación genómica antes del 2014 (Hamshere y cols., 2013; Ikeda y cols., 2011; Ma y cols., 2011; Irish Schizophrenia Genomics Consortium y Wellcome Trust Case Control Consortium 2, 2012; Lencz y cols.; 2013; Rietschel y cols., 2012; Ripke y cols., 2013a; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011; Steinberg y cols., 2011a; Yamada y cols., 2011). Entre dichos estudios, el más impactante fue el de Ripke y cols. (2013a), en el cual 22 *loci* no solapantes mostraron asociación con significación genómica, y 13 de ellos no se habían encontrado hasta entonces implicados en estudios de asociación. A raíz de este estudio, que utilizó más de 21000 casos y 38000 controles, superando los tamaños de muestra de otros estudios hasta entonces, se conoció la implicación del microRNA *MIR137* y se vio que las señales de asociación estaban enriquecidas en las funciones de señalización del calcio. Además, un gran número de los genes implicados son dianas en la regulación post transcripcional de *MIR137*, un hecho también extensible a los factores de transcripción asociados *TCF4* y *ZNF804A*, lo que sugirió la posibilidad de que la regulación de la expresión génica, mediante esta y otras vías, jugase un papel muy importante en la esquizofrenia.

En coherencia con los resultados del trabajo de Purcell y cols. (2009), en 2011 se realizó un GWAS con muestras conjuntas de esquizofrenia y trastorno bipolar que reveló, de nuevo, la implicación de varios *loci* en ambos trastornos (Sklar y cols., 2011). Este solapamiento entre ambas enfermedades ya había quedado de manifiesto en los genes *ZNF804A* (Donovan y cols., 2008) y en *CACNA1C* (Green y cols., 2010)

A este trabajo le siguieron otros que estudiaron el posible solapamiento entre esquizofrenia, trastorno bipolar, autismo, trastorno por déficit de atención e hiperactividad y depresión mayor (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013a y 2013b). De nuevo se observó solapamiento en el conjunto de enfermedades, centrado sobre todo en variantes en canales de Calcio, revelando una posible susceptibilidad genética indiferenciada, y ahondando todavía más en el concepto de enfermedad psiquiátrica como síndrome no bien delimitado. No obstante, el mayor solapamiento de variación genética común se observó entre los trastornos de inicio en edad adulta: esquizofrenia, depresión y trastorno bipolar; mientras que en menor grado entre autismo y esquizofrenia.

El mayor estudio de GWAS hasta la fecha es el realizado hace poco más de un año (Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC), 2014). Fue posible gracias a la utilización de 49 cohortes independientes que englobaron 34241 casos y 45604 controles, además de 1235 tríos, lo cual doblaba los número de estudio más grande hasta entonces (Ripke y cols., 2013a). Además, los SNP que pasasen un umbral de significación de $p < 10^{-6}$, se sometieron a un análisis independiente con 1513 casos y 66236 controles de *DECODE genetics*. Los grandes tamaños muestrales se tradujeron en resultados fructíferos, encontrándose 128 señales de asociación estadísticamente independientes en 108 *loci* no solapantes asociados a esquizofrenia. Tal número de asociaciones hizo posible, por fin, obtener una idea más consistente de las rutas biológicas existentes detrás de la variación genética común, y que se describen a continuación.

- En primer lugar, este es el primer estudio en el que se obtiene una evidencia de asociación con la ruta dopaminérgica a raíz de un estudio GWAS, lo cual vincula, por primera vez y en un estudio libre de hipótesis previas, a la principal ruta de dianas farmacológicas en esquizofrenia con la genética de la enfermedad. El gen asociado es el receptor de dopamina D2 (*DRD2*) y, pese a que fueron necesarias miles de muestras para encontrar una asociación significativa pero de efecto menor, esto demuestra que incluso el hallazgo de asociaciones con efecto limitado sobre la enfermedad puede tener repercusiones muy trascendentes desde el punto de vista del tratamiento farmacológico.
- En línea con los estudios de GWAS inmediatamente anteriores (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013b; Ripke y cols., 2013a), las principales señales de asociación implican a la ruta glutamatérgica, con asociaciones en los genes *GRI1A1*, *GRIN2A* y *GRM3*, a miembros de los canales iónicos del Calcio como *CACNA1C*, *CACNA1I* o *CACNB2*, y a genes con relevancia en plasticidad sináptica. Esto consolida la hipótesis previa según la cual la disrupción de la vía del glutamato y de la homeostasis neuronal del Calcio son factores de predisposición claves en el desarrollo de la esquizofrenia.
- La existencia del componente poligénico fue completamente demostrada. Para un prevalencia del 1% e independencia de los SNP analizados, las estimaciones de heredabilidad subieron de un 3% a un 7%, por lo que la varianza explicada era mayor, aunque seguía estando alejada de los valores esperados en la enfermedad.
- El solapamiento con otras enfermedades psiquiátricas ya había quedado demostrado con la explicación de forma significativa de una parte de la varianza genética del trastorno bipolar con el componente poligénico de la esquizofrenia, algo que no sucedía con autismo. Sin embargo, en este estudio se observó que los *loci* asociados estaban enriquecidos en genes con variantes *de novo* no sinónimas no solo en esquizofrenia,

sino también en autismo y discapacidad intelectual con lo que el espectro de solapamiento aumentaba respecto a anteriores evidencias.

- La mayor limitación del GWAS fue la ausencia de respuestas para muchas de las asociaciones encontradas, donde no se encontró una clara funcionalidad afectada. Prueba de ello es que solo 10 de las señales de asociación se pueden explicar por cambios no sinónimos en la secuencia proteica y solo 12 fueron eQTLs conocidos. De hecho, el SNP no sinónimo más significativo en esquizofrenia ya fue encontrado con anterioridad a este estudio (Carrera y cols., 2012).
- Por último, y como consecuencia de lo anterior, el PGC hizo uso de datos *ómicos* de otros dos grandes proyectos centrados en la regulación de la expresión y las modificaciones epigenéticas del genoma, como son el proyecto ENCODE (The Encyclopedia of DNA Elements) y el proyecto Roadmap (ENCODE Project Consortium, 2012; Kundaje y cols., 2015), que se explican detalladamente en el último bloque de la introducción. La utilización de marcas de acetilación de histonas para definir elementos potenciadores de la expresión o *enhancers* reveló que existe un enriquecimiento significativo de las señales de asociación en regiones reguladoras específicas de cerebro adulto y de células implicadas en la respuesta inmune, éstas últimas, incluso, tras eliminar aquellas con marcada actividad cerebral y del MHC. Por tanto, este estudio consolida el papel del sistema inmune desde un punto de partida genético y de la regulación genética como factores determinantes en el desarrollo de la esquizofrenia.

Muy recientemente, el *Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium*, un consorcio que engloba a cientos de colaboradores, publicó un estudio en el que se hace una interpretación funcional de las señales de asociación de los GWAS de distintos trastornos psiquiátricos en base a las bases de datos de rutas biológicas o *pathways* existentes demostró que el sistema inmune, la sinapsis neuronal y sobre todo, la metilación de histonas, constituían las asociaciones más significativas compartidas en esquizofrenia, trastorno bipolar y depresión mayor, sugiriendo un papel crucial de la regulación de la expresión génica en la patología psiquiátrica (Network, T., & Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015).

1.2.2.4. Detección de Variaciones en el número de copias (CNV)

El avance en el desarrollo de las tecnologías de genotipado, además de permitir la realización de GWAS a precios razonables, representó una posibilidad para analizar a gran

escala variaciones estructurales a lo largo del genoma, y en particular, variaciones en el número de copias (del inglés *Copy Number Variant (CNV)*).

Las CNV son un tipo de variantes genéticas estructurales, generalmente deleciones o duplicaciones de un segmento de ADN, que afectan al número de copias del segmento y la dosis génica de los genes incluidos en ella. El tamaño de la región afectada es, en la mayor parte de los casos, submicroscópico (< 500 kb), de modo que resulta muy complicada su detección por métodos de cariotipado clásicos. Por ello, el estudio de CNV se vio favorecido por el desarrollo de técnicas como la *hibridación genómica comparada* (CGH) o *microarrays* de SNP similares a los utilizados en GWAS, donde la resolución es un parámetro dependiente de la plataforma de genotipado utilizada. Otras estrategias para su detección, como la secuenciación, resultan cada vez más asequibles dada la reducción de costes que ha venido experimentando esta tecnología a lo largo de los últimos años. De hecho, la secuenciación es la estrategia con una resolución máxima para detectar variaciones estructurales muy pequeñas, difíciles de detectar incluso con métodos de genotipado de alta resolución (Mills y cols., 2011).

En 2004, dos estudios demostraban que las CNV se encontraban ampliamente distribuidas por todo el genoma y representaban una importante fuente de variación genética humana (Iafrate y cols., 2004; Sebat y cols., 2004), del mismo modo que lo son las variantes puntuales como los SNP. En promedio, se estima que existen más de 1000 CNV en el genoma implicando un número aproximado de 4×10^6 bases nucleotídicas debido a ellas (Conrad y cols., 2010; Mills y cols., 2011). Esto supone que, pese a que el número de SNP es de tres órdenes de magnitud por encima del de CNV, las variaciones estructurales ocurren en aproximadamente un 5% del genoma y suponen una contribución a la variabilidad genética en la población del 0.1%, que se suma al 0.1% de la diferencia existente en la secuencia de ADN a nivel de nucleótido.

La formación de CNV es un fenómeno, por tanto, bastante frecuente y que se puede deber a varios mecanismos ampliamente descritos. La *recombinación alélica no homóloga* (Hurles y Lupski, 2006) fue el mecanismo propuesto mayoritariamente. Según este, la alteración estructural se produce principalmente durante la meiosis, donde tiene lugar un entrecruzamiento erróneo entre cromosomas homólogos o cromátidas hermanas, inducido por la presencia de elementos repetitivos como los *low copy repeats* (LCRs). No obstante, otros mecanismos también explicarían la aparición de CNV, como la retrotransposición (Lupski, 2010), el mecanismo de *fork stalling* (Lee, Carvalho y Lupski, 2007) o el de *non homologous End-joining* (Lieber, 2010).

La contribución de las CNV a la variabilidad genética humana es manifiesta y, por ello, muy probable su implicación en enfermedades de base genética así como un mecanismo muy

importante en procesos evolutivos (Nguyen y cols., 2008; Conrad y cols., 2010). Dada la recurrencia de esta forma de variación genética estructural, dos modelos surgieron para explicar su papel en las enfermedades complejas. Un modelo basado en polimorfismos de número de copias (del inglés *copy number polymorphism (CNP)*), donde las variantes presentan un rango en el número de copias de 0 a 30 y una frecuencia superior al 1%, se contrapone a otro basado en la implicación de CNV individuales o de baja frecuencia ($< 1\%$), que existen en forma de pocas copias o deleciones, de alta penetrancia, sometidas a una alta presión purificadora y con una alta tasa de reposición por eventos *de novo* (Girirajan y cols., 2011).

En el contexto de las enfermedades psiquiátricas, las CNV se han visto implicadas en forma de variantes raras (Sebat y cols., 2009). El foco en las CNV en el ámbito de estas enfermedades tiene su precedente en el caso del autismo, donde estudios citogenéticos reportaron la presencia de alteraciones estructurales en un 6-7% de los casos (Folstein y Rosen-Sheidley, 2001). Por otro lado, antes de la puesta a punto de las técnicas de hibridación o de los arrays de SNP, ya se conocía la incidencia de la duplicación del locus 15q11-13 en algunos casos de autismo idiopático (Gillberg, 1998) así como la microdelección del locus 22q11.2 en esquizofrenia (Karayiorgou y cols., 1995). Esta última variante tiene un origen *de novo* en más de un 70% de los casos (Ryan y cols., 1997) y es la causante del síndrome velocardiofacial (Kobrynski y Sullivan, 2007), una patología con fenotipo muy notable y de manifestaciones variables, entre las cuales destacan las alteraciones cardiovasculares, la inmunodeficiencia, retraso mental, alteraciones cognitivas y, en un 30% de los enfermos, psicosis predominantemente en forma de esquizofrenia (Murphy y cols., 1999). Este hecho convierte a la delección 22q11.2 en la variante de mayor riesgo para padecer esquizofrenia (con un OR mayor de 25) y donde, además, no se ha reportado ningún portador sin síntomas clínicos asociados a la CNV (Malhotra y Sebat, 2012). Esta CNV afecta a un número de genes que oscila entre 30 y 60 dependiendo del tamaño de la misma, y la alteración de la dosis génica de alguno de ellos podría ser la razón de la relación con esquizofrenia. Entre los más importantes candidatos figuran *COMT*, por su papel en la degradación de la dopamina, *DGCR8*, cuya delección conlleva déficits en memoria y aprendizaje (Ouchi y cols., 2013) mediante alteraciones en los niveles de microRNA, algo que ya se ha encontrado anteriormente en pacientes con esquizofrenia (Beveridge y cols., 2010; Perkins y cols., 2007). De igual modo, otros genes se han visto implicados en diversos estudios como *PRODH* (Bender y cols., 2005), *GNBIL* (Ishiguro y cols., 2008) (Chun y cols., 2014) o *ZDHHC8* (Ota y cols., 2013).

Los estudios de asociación han demostrado que las CNV de riesgo en esquizofrenia son generalmente grandes (> 100 kb) y de frecuencia menor al 1%. El primer estudio caso - control a gran escala llevado a cabo para investigar la incidencia de CNV en esquizofrenia encontró tres veces más CNV raras en pacientes que en controles (Walsh y cols., 2008). Estos hallazgos

fueron corroborados en un estudio más largo de 3391 y 3181 casos y controles, respectivamente, donde se observó un incremento del 15% en la incidencia de CNV raras en esquizofrénicos (The International Schizophrenia Consortium, 2008). Aunque menor que el anterior, este incremento se hacía más pronunciado cuanto más infrecuente (o incluso individual) era la variante y mayor contando solo las que contenían genes. Stefansson y cols., además, encontraron asociaciones significativas de CNV *de novo* en los *loci* 1q21.1, 15q11.2 y 15q13.3, convirtiéndose en el primer estudio que encontraba asociaciones en CNV concretas (Stefansson y cols., 2008). De igual modo, otros estudios encontraron resultados similares (Ikeda y cols., 2010; Kirov y cols., 2009a; Levinson y cols., 2011; Need y cols., 2009), y en el último metaanálisis realizado, en el cual el genotipado de 6882 casos y 6316 controles se suman a los de estudios anteriores, se refuerzan los resultados previamente publicados (Rees y cols., 2014). Un 2.5% de los casos presentan CNV raras frente a un 0.9% de controles, siendo significativo el incremento en 15 *loci* (**tabla 1.3**). La asociación de las CNV raras es, además, debida a eventos recurrentes, ocurriendo en regiones con una tasa de mutación elevada.

Una posible explicación de la alta incidencia de CNV es la posibilidad de que en gran medida surjan como eventos *de novo*. Esta hipótesis fue estudiada por Xu y cols. (2008) mediante análisis de tríos, y demostraron que el incremento encontrado en CNV raras en casos respecto a controles era debido en mayor medida a variantes esporádicas ($P = 0.00078$) que a heredadas ($P = 0.049$), siendo las primeras ocho veces más frecuentes en esquizofrenia. Estudios posteriores demostraron, de nuevo, una mayor incidencia de CNV *de novo* en esquizofrenia (Kirov y cols., 2012; Malhotra y cols., 2011).

El estudio de Walsh y cols (2008) reveló que los genes afectados por las deleciones y duplicaciones en CNV raras estaban enriquecidos en categorías funcionales relacionadas con la sinapsis y el neurodesarrollo, lo que supuso la mejor demostración de la hipótesis del neurodesarrollo en esquizofrenia hasta el momento. Posteriormente, analizando más detalladamente las implicaciones funcionales de los genes de CNV *de novo* en esquizofrenia se vio que existía un enriquecimiento en las proteínas del receptor de glutamato NMDAR y proteínas ARC relativas a la densidad postsináptica implicadas en funciones relacionadas con el citoesqueleto celular (Kirov y cols., 2012). Esta implicación en el neurodesarrollo cerebral es consistente con el solapamiento que existe, respecto a estas variantes, con otros trastornos neuropsiquiátricos y del desarrollo. La variabilidad fenotípica asociada a las CNV debe estar regulada por otros aspectos genéticos, incluidas variantes raras o variación común, y factores ambientales, pudiendo estar mediados por mecanismos epigenéticos.

Tabla 1. 3. Principales CNVs asociadas significativamente a esquizofrenia (Rees y cols., 2014). Evidencias de asociaciones de las mismas CNVs con otros trastornos psiquiátricos se recogen del estudio de Malhotra y cols., (2012) y las abreviaciones de la tabla representan las siguientes enfermedades: RM/MC = Retraso mental o malformación congénita; ASD = Trastorno del espectro autista; DB = Trastorno bipolar; MDD = Depresión mayor; EP = Epilepsia; ADHD = Trastorno por déficit de atención e hiperactividad; OCD = Trastorno obsesivo compulsivo; ANX = Otros trastornos de ansiedad. En tono gris se marca la microduplicación de *VIPR2*, sin significación estadística en el estudio referido de la tabla, pero sí en varios estudios anteriores. En las CNVs de los *loci* 7q11.23 y 15q11.2 - 13.1 se incluyen, entre paréntesis, los síndromes mendelianos que causan la duplicación y delección en respectivas regiones y que son el síndrome de Williams-Beuren (WBS) y el síndrome de Prader - Willi (PWS), respectivamente.

CNV	Tipo	Posición (Mb)	N SCZ	N CO	OR	(95 % CI)	P-valor	Otras asociaciones (Malhotra y cols., 2012; Rees y cols., 2014; Doherty y Owen; 2014)
1q21.1	del	chr1:146.57-147.39	33/19056	17/81821	8.35	(4.65-14.99)	4.1 x 10 ⁻¹³	RM/MC, ASD, ADHD
1q21.1	dup	chr1:146.57-147.39	21/16247	24/64046	3.45	(1.92-6.20)	9.9 x 10 ⁻⁵	RM/MC, ASD, BD, MDD
2p16.3 (<i>NRXN1</i>)	del	chr2:50.15-51.26	33/18762	10/51161	9.01	(4.44-18.29)	1.6 x 10 ⁻¹¹	RM/MC, ASD
3q29	del	chr3:195.73-197.34	14/17005	1/69965	57.65	(7.58-438.44)	1.5 x 10 ⁻⁹	RM/MC, BD
7q11.2 (<i>WBS</i>)	dup	chr7:72.74-74.14	14/21269	2/34455	11.35	(2.58-49.93)	6.9 x 10 ⁻⁵	RM/MC, ASD, ADHD, ANX
7q36.3 (<i>VIPR2</i>)	dup	chr7:158.82-158.94	15/14218	17/24815	1.54	(0.77-3.09)	0.27	
15q11.2	del	chr15:22.80-23.09	116/19547	227/81802	2.15	(1.71-2.68)	2.5 x 10 ⁻¹⁰	RM/MC, EP, ADHD, OCD
15q11.2 - 13.1 (<i>PWS</i>)	dup	chr15:24.82-28.43	12/14464	3/47686	13.2	(3.72-46.77)	5.6 x 10 ⁻⁶	RM/MC, ASD
15q13.3	del	chr15:31.13-32.48	26/18571	15/80422	7.52	(3.98-14.19)	4 x 10 ⁻¹⁰	RM/MC, ASD, EP, ADHD
16p13.11	dup	chr16:15.51-16.30	37/12029	93/69289	2.3	(1.57-3.36)	5.7 x 10 ⁻⁵	RM/MC, ADHD
16p11.2	del distal	chr16:28.82-29.05	13/20732	5/27045	3.39	(1.21-9.52)	0.017	RM/MC, ASD
16p11.2	dup	chr16:29.64-30.20	58/16772	19/63068	11.52	(6.86-19.34)	2.9 x 10 ⁻²⁴	RM/MC, ASD, BD, MDD, ADHD, ANX
17p12	del	chr17:14.16-15.43	12/12773	17/65402	3.62	(1.73-7.57)	0.0012	
17q12	del	chr17:34.81-36.20	5/14024	4/74447	6.64	(1.78-24.72)	0.0072	RM/MC, ASD
22q11.2	del	chr22:19.02-20.26	56/19084	0/77055	NA	(28.27-	4.4 x 10 ⁻⁴⁰	RM/MC, ASD, BD, MDD, ADHD, ANX

Una posibilidad es que las CNV provoquen un riesgo no específico a trastornos del neurodesarrollo que se ve modulado por la presencia de otro factor adicional. Este hecho se conoce como la hipótesis de doble impacto. Un ejemplo es el estudio llevado a cabo por Girirajan y cols. (2010), en el cual se demostraba que el riesgo de la microdelección de 520 kb en el locus 16p12.1 para padecer retraso en el desarrollo era más significativo en presencia de una segunda CNV adicional en los casos. En esta línea, se ha considerado también el hecho de que las CNV sean factores de riesgo asociado a retraso mental y trastornos del desarrollo severos que, en conjunción con otras variantes genéticas o factores ambientales, puedan desembocar en una amalgama de posibilidades donde predominan aquellas con un claro componente de neurodesarrollo (Malhotra y cols., 2012). Esto explicaría el hecho de que prácticamente todas las CNV asociadas a esquizofrenia presentan también riesgo a trastornos de desarrollo y autismo, y otras que no presentan significación en esquizofrenia sí la tengan en trastornos del desarrollo más severos como los anteriores (Malhotra y cols., 2012).

Las asociaciones de las CNV con la esquizofrenia y otros trastornos del desarrollo se han constatado, también, a través de estudios animales. Muchos trastornos del comportamiento pueden provocarse mediante modelos animales de estas variantes. Por ejemplo, ratones con la duplicación 15q11-3 de herencia paterna sufren ansiedad, poca flexibilidad en el comportamiento y peor interacción social (Nakatani y cols., 2009). Alteraciones en el comportamiento similares o variadas también ocurren en otros casos, como en los modelos de duplicaciones y deleciones del locus 16p11.2 (Horev y cols., 2011) y deleciones de *NRXN1* (Etherton y cols., 2011) y *CNTNAP2* (Peñagarikano y cols., 2009). A estas alteraciones en el comportamiento también se le suman otras en la neuroanatomía cerebral, más extrapolable que los estudios de comportamiento, como la reducción y el incremento del volumen cerebral asociado a la respectiva duplicación o delección del locus 16p11.2 (Horev y cols., 2011) y otras alteraciones en las funciones neuronales como la liberación de neurotransmisores o el funcionamiento de los canales de calcio en *NRXN1* (Missler y cols., 2003). En el caso de la CNV 16p11.2, ensayos en pez cebra han demostrado que el gen *KCTD13* es el principal responsable de la macrocefalia o microcefalia asociada a los fenotipos relativos a la delección o duplicación en esta región, respectivamente, pudiendo estar implicados también los genes *MVP* y *MAPK3* (Golzio y cols., 2012).

En conjunto, se pueden extraer varias conclusiones de los estudios de CNV. En primer lugar, sirven como demostración empírica de que parte de la varianza genética de la esquizofrenia tiene una explicación debida a variación rara, no considerada en los estudios de GWAS. Estos estudios revelan que la predisposición a sufrir esquizofrenia mediada por esta variación rara tiene un impacto importante en genes relacionados con los procesos neuronales del desarrollo cerebral, hecho que consolida de forma consistente las bases de la hipótesis del

neurodesarrollo. Existe un solapamiento en la funcionalidad alterada por la variación común, descrita en los GWAS, y la causada por CNV. En ambos casos queda demostrado que la ruta glutamatérgica tiene un papel esencial, principalmente a través del receptor NMDAR; y de igual modo, la densidad post-sináptica. Sin embargo, una importante diferencia radica en el hecho de que, si bien los GWAS encontraban un mayor solapamiento de la esquizofrenia con trastornos psiquiátricos propios de la adultez (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013a; Purcell y cols., 2009), en el caso de las CNV, el mayor solapamiento tiene lugar con trastornos que implican alteraciones durante el neurodesarrollo (Malhotra y cols., 2012; Rees y cols., 2014; Walsh y cols., 2008). No obstante, esto, lejos de tener que ser un problema de entendimiento, puede ser visto como una demostración de la complejidad del espectro esquizofrénico, encontrándose a caballo entre el neurodesarrollo y los trastornos adultos; y una muestra del continuo que representa la patología psiquiátrica. En cualquier caso, existen evidencias de la implicación de un mismo locus en esquizofrenia mediante variantes raras o comunes. Tal es el caso del gen TAOK2, presente en la región de la CNV 16p11.2 y con un SNP también asociado a esquizofrenia (Steinberg y cols., 2014), o del gen ZNF804A, implicado en el riesgo a la enfermedad por varios estudios GWAS y en el cual también hallado una CNV de riesgo (Steinberg y cols., 2011b). Por tanto, la búsqueda de variantes genéticas de riesgo en genes candidatos por formar parte de CNV asociadas es una estrategia acertada tanto para localizar los genes implicados en esos *loci* como para revelar variabilidad de riesgo asociada a ellos.

1.2.2.5. Next Generation Sequencing (NGS) y estudios de secuenciación en esquizofrenia.

Los estudios de secuenciación de interés en esquizofrenia son muy recientes. La elevada heterogeneidad genética de la enfermedad y la consiguiente inviabilidad económica para la utilización de estas técnicas, hizo esperar la llegada de los primeros resultados de secuenciación en esquizofrenia.

La historia de la secuenciación de ADN tiene su punto de partida en los métodos de Gilbert (Maxam y Gilbert, 1977) y de Sanger (Sanger y cols., 1977). Este último representó una notable mejoría respecto al primero por la menor utilización de compuestos tóxicos y radioisótopos durante el proceso, lo que hizo que fuese la técnica elegida para los estudios de secuenciación durante los siguientes treinta años y la base del Proyecto Genoma Humano (PGH), que finalizó, tras más de diez años de trabajo, la secuencia final del genoma (International Human Genome Sequencing Consortium, 2004). Más allá de los obvios y comentados avances de este hecho, y del impulso que esto proporcionó a los estudios genéticos en enfermedades complejas, las dificultades encontradas durante el proceso y la enorme

inversión requerida pusieron de manifiesto la necesidad de mejorar las técnicas de secuenciación. Por esa razón, el Instituto nacional de investigación genómica humana de EEUU (NHGRI) financió el desarrollo de una metodología avanzada y se planteó la meta de conseguir, en un plazo de diez años desde entonces, la secuenciación del genoma humano con un coste tope de 1000\$ (Liu y cols., 2012; van Dijk y cols., 2014). Esto supuso la aparición de las técnicas de *ultrasecuenciación* o *secuenciación de nueva generación*, del inglés *Next Generation Sequencing (NGS)*, quedando la de Sanger como la de primera generación. Las principales ventajas que supuso esta nueva tecnología de secuenciación fueron las siguientes:

- El proceso de clonado en bacterias fue sustituido por la creación de librerías, con diferencias notables entre cada plataforma (Mardis, 2008).
- La mejoría técnica tuvo como principal consecuencia el pasar de pocos cientos a miles, e incluso millones, de reacciones de secuenciación en paralelo, aumentando el rendimiento de forma notable.
- Ya no se necesitó la detección electroforética del producto de secuenciación, simplificando y acelerando el proceso de forma significativa.

Estas ventajas generaron, por contrapartida, una enorme necesidad de renovar y mejorar los métodos de alineamiento del genoma, dada la generación de miles de lecturas o *reads*. La primera plataforma en salir al mercado fue la de pirosecuenciación **454** de Roche (2005), y de forma consecutiva salieron la primera plataforma de **Illumina** (2006) y **SOLiD** (2007) de Applied Biosystems (hoy Life Technologies). La química de cada proceso es diferente y en ello radican las diferencias en tiempo, rendimiento y aplicaciones preferentes que se recogen en la **tabla 1.4**. Más adelante, en 2010, salió el **Personal Genome Machine (PGM)** de Ion Torrent (hoy también parte de Life Technologies), de los mismos creadores del 454, y con una química parecida pero en el que destaca el uso de un semiconductor como método de detección en lugar de la detección óptica por fluorescencia. Finalmente, también en 2010, salió al mercado la primera y más importante versión de una nueva metodología de ultrasecuenciación que, por su novedosa alternativa, se denominó *secuenciación de tercera generación* (Schadt y cols., 2010). Es el caso del **Pac Bio RS** de Pacific Biosciences, que se caracteriza por sus *reads* de gran tamaño, ideales para el ensamblaje de genomas *de novo* o mejorar los ya existentes, y su impresionante rapidez en comparación con Illumina y SOLiD, las plataformas ampliamente más utilizadas. No obstante, su bajo rendimiento y el alto coste de utilización de esta plataforma hace todavía cuestionable su utilización preferente en la actualidad.

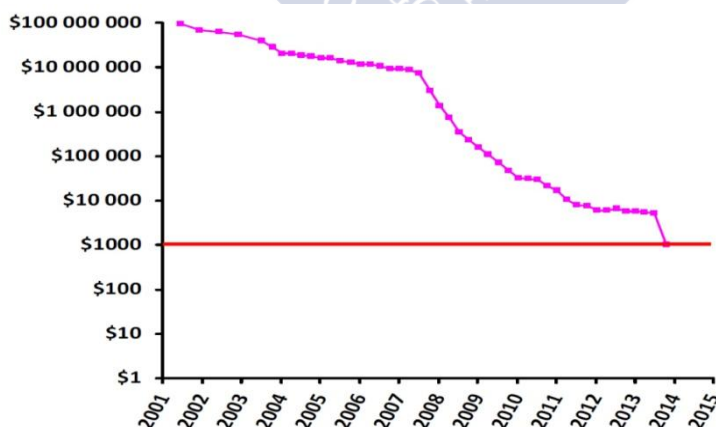
En el ámbito económico, se podría decir que Illumina ha sido la primera tecnología en llegar al objetivo de 1000\$ por genoma secuenciado (**figura 1.6**), aunque esto ha sido cuestionado porque para alcanzar ese coste implicaría la difícil rutina de secuenciación masiva

de 18000 genomas al año (van Dijk y cols., 2014). En cualquier caso, la reducción en los costes de secuenciación ha sido la principal ventaja de la NGS y la razón de que se empezase a utilizar de forma cada vez más numerosa en el estudio genético y diagnóstico de enfermedades mendelianas y complejas.

Tabla 1. 4. Principales características, ventajas y desventajas de las plataformas de secuenciación más importantes (Liu y cols., 2012; van Dijk y cols., 2014).

	454	Illumina	SOLiD	Ion Torrent	Pac Bio
Longitud de lectura (pb)	110 - 1000	35 - 300	35 - 75	200 - 400	4000 - 20000
Rendimiento máximo (Gb)	700	1800	320	10	0.5
Tiempo	23 hrs	3-10 días	7 - 14 días	pocas horas	Minutos - horas
tasa de error	alta	media	baja	alta	alta
Principales ventajas	- Rápido - Lecturas largas	- Alto rendimiento	-Alta precisión	- Muy rápido	- Lecturas muy largas (ensamblado <i>de novo</i>) - Muy rápido
Desventajas	- Reactivos caros - Poco preciso - Cierre de plataforma	- Manejo complicado - Lecturas cortas	- Lecturas cortas - Lento - Cierre de plataforma	- Poco preciso	- Poco preciso - Rendimiento bajo

Figura 1. 6. Reducción en el coste de secuenciación de un genoma humano a lo largo de los años (Extraída de van Dijk y cols., 2014).



La ultrasecuenciación ha supuesto una revolución en el ámbito de la genética, posibilitando no solamente la secuenciación del genoma a un precio razonable, sino una infinidad de aplicaciones derivadas. Ejemplo de ellas son la secuenciación de RNA (RNA-seq),

que ha demostrado una mayor sensibilidad que los microarrays de expresión y permite el estudio de la variación genética a nivel de transcriptoma, así como los mecanismos de *splicing* que ocurren en los distintos tipos celulares y tejidos de interés (Wang y cols., 2009). Otro ejemplo es el mapeo de las interacciones ADN - proteína, así como de marcas epigenéticas como modificaciones histónicas, mediante técnicas de *inmunoprecipitación de la cromatina seguidas de secuenciación (CHIP-Seq)*, y que han hecho posible el estudio de la regulación del genoma de una forma sin precedentes (ENCODE Project Consortium, 2012; Kundaje y cols., 2015).

Por otra parte, la secuenciación ha permitido la creación de grandes consorcios cuyo objetivo fuese generar repositorios de variantes a lo largo del genoma de miles de individuos. Tal es el caso del proyecto 1000 genomas (1000 Genomes Project Consortium, 2010; 1000 Genomes Project Consortium, 2012), con la secuenciación inicial de 1167 individuos en una primera fase, 633 en una segunda y 700 en la final, con los que se pretende obtener un mapa de la variación genética existente en el genoma humano a través de muchas poblaciones. Esto supone una base de datos muy útil para comparar las frecuencias alélicas entre individuos con un fenotipo concreto y estas poblaciones control. De igual modo, existen otros proyectos de secuenciación masiva restringidos al exoma, como el caso del *NHLBI Exome Sequencing Project* (Fu y cols., 2013) o del *Exome Aggregation Consortium* (Lek y cols., 2015)

En el contexto de la esquizofrenia, y tras el último estudio de GWAS con decenas de miles de casos y controles, ha quedado demostrada la contribución de la variación común a la esquizofrenia. No obstante, el componente poligénico del 7% no explica, ni de lejos, la enorme heredabilidad estimada para este trastorno. Además, generalmente, esa variación no ha podido ser asignada a *loci* o genes concretos debida al bajo riesgo asociado a gran parte de las variantes comunes que componen esa susceptibilidad. Por otro lado, dadas las evidencias encontradas acerca de la contribución de las variantes raras mediante el estudio de CNV, un lógico paso a seguir sería explorar el escenario de la variación rara en su conjunto, estudiando las mutaciones puntuales, así como pequeñas inserciones y deleciones, a lo largo del genoma. Los estudios de asociación GWAS, basados en la estrategia indirecta de DL, no tienen capacidad para cubrir ese espectro de variación rara y, por ello, los avances en secuenciación y la aparición de las tecnologías de NGS hicieron muy viable el salto a este nuevo escenario en el que abordar el estudio de variantes raras no estructurales de predisposición a esquizofrenia.

Pese a la reducción de los costes operativos de las técnicas de ultrasecuenciación desde su eclosión, éstos todavía se sitúan muy por encima de los de las plataformas de genotipado. Esto provocó que gran parte de los estudios llevados a cabo en enfermedades complejas como la esquizofrenia siguiesen estrategias de selección de las regiones de interés. Una primera

aproximación en este sentido sería abordar el estudio del total del exoma, es decir, la parte del genoma codificante (Hodges y cols., 2007). El exoma representa menos del 2% del genoma pero alberga un 85% de las variantes patogénicas ligadas a las enfermedades mendelianas, por lo que representa una alternativa económica para capturar regiones que, a priori, podrían ser más susceptibles de albergar variantes patogénicas (Choi y cols., 2009b). De este modo, frente a la secuenciación de genoma completo (WGS, del inglés *Whole genome sequencing*), una alternativa económica muy viable es la secuenciación del exoma (WES, del inglés *Whole exome sequencing*). Una alternativa todavía más viable económicamente es la resecuenciación dirigida (o *target resequencing*) de regiones específicas amplificadas mediante PCR, que está muy indicado cuando hay algunos *loci* que interesa estudiar en particular, consiguiendo una mayor cobertura y sin los elevados filtros de calidad implementados de forma sistemática en los estudios de exoma (Rehm, 2013).

En esquizofrenia, los primeros estudios, realizados mediante secuenciación de Sanger, se centraron en regiones específicas, como el gen *DISC1* (Song y cols., 2008), el gen *ABCA13* (Dwyer y cols., 2011) u otras regiones ligados al cromosoma X, estudiados debido a la mayor incidencia de esquizofrenia en hombres (Piton y cols., 2011). Posteriormente, Crowley y cols. (2013) secuenciaron diez genes clásicos con evidencias previas en esquizofrenia (*COMT*, *DAOA*, *DISC1*, *DRD2*, *DRD3*, *DTNBP1*, *HTR2A*, *NRG1*, *SLC6A3* y *SLC6A4*) en 727 casos y 733 controles, genotipando las variantes encontradas en muestras independientes de mayor tamaño. Los resultados de este estudio, aunque en algún caso sirvieron para poner de manifiesto la mayor presencia de variantes raras en casos respecto a controles, no lograron detectar ninguna variante que confiriese alto riesgo para padecer esquizofrenia.

Los primeros estudios de exoma, realizados mediante NGS, tampoco resultaron mucho más fructíferos (Need y cols., 2012; Timm y cols., 2013), resaltando únicamente el hallazgo de variantes raras no sinónimas en tres genes relacionados con el receptor de glutamato NMDAR a raíz de la secuenciación de exomas de cinco familias (Timm y cols., 2013). Sin embargo, resultaron clarificadores para descartar la existencia de variantes exónicas a frecuencias entre 1 - 5% y OR asociado mayor que 2, delimitando la arquitectura genética de la esquizofrenia (Need y cols., 2012). Al no encontrarse variantes de riesgo elevado, estos estudios pusieron de manifiesto la necesidad imperiosa de conseguir tamaños muestrales mayores, analizar la acumulación de variantes en genes o grupos de genes con funciones relacionadas a fin de alcanzar una mayor potencia estadística y explorar nuevos escenarios como el de regiones reguladora no codificantes.

Por otra parte, la búsqueda de variantes raras *de novo*, tal y como había sucedido en el estudio de CNV, demostró que las mutaciones espontáneas a nivel de nucleótido también juegan

un papel muy importante en la susceptibilidad a padecer esquizofrenia, estando sobrerrepresentadas en individuos afectos (Girard y cols., 2011; Gulsuner y cols., 2013; Xu y cols., 2012a). Además, en relación a los genes encontrados con variantes *de novo*, se encontró un enriquecimiento en aquellos altamente expresados en etapas fetales (Xu y cols., 2012a) y, específicamente, en el córtex prefrontal (Gulsuner y cols., 2013). Gulsuner y cols demostraron que la edad paterna es un factor significativo para el número de mutaciones *de novo* en la progeñie y, por ende, un factor de riesgo a tener en cuenta, algo en la línea de evidencias anteriores (Kong y cols., 2012). En este sentido, sin embargo, estudios recientes han encontrado tanto un riesgo asociado a la mayor edad paterna como a menor edad materna, generando un escenario de interacción más complejo de interpretar (McGrath y cols., 2014). Lo que sí es un hecho claro es que, del mismo modo que en el caso de CNV (Walsh y cols., 2008; Xu y cols., 2008), estos estudios también apuntan a una susceptibilidad causada por eventos raros *de novo* que interrumpen la función de algunos genes como *LAMA2* o *DPYD*, importantes durante el neurodesarrollo (Sullivan, Kendler y Neale, 2003; Weinberger, 1986).

Los dos estudios más importantes de secuenciación de exomas en esquizofrenia fueron publicados en el 2014 (Fromer y cols., 2014; Purcell y cols., 2014). El trabajo de Fromer y cols., el estudio de variación *de novo* en esquizofrenia más grande hasta la fecha con un total de 623 tríos utilizados, no replicó una mayor incidencia de variantes *de novo* en casos respecto a controles, ni de forma general ni conteniendo únicamente las mutaciones de pérdida de función o *Loss of function (LoF)*, definidas como aquellas más patogénicas por ser *non sense* (que provoca la aparición de un codón de terminación prematuro), de *splicing* (en los sitios de corte y empalme del ARN) o *frameshift* (que altera la pauta de lectura). Sin embargo, sí se observó que las variantes no sinónimas aparecían sobrerrepresentadas significativamente en esquizofrénicos en sets de genes de las proteínas glutamatérgicas de la densidad postsináptica del complejo ARC y del receptor de glutamato NMDAR, previamente implicados en esquizofrenia (Kirov y cols., 2012; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Timm y cols., 2013), y sitios de unión a ARN mensajero de FMRP, un hecho recientemente observado en genes de CNV en esquizofrenia (Szatkiewicz y cols., 2014).

FMRP (del inglés *fragile X mental retardation protein*) es una proteína de unión a ARN cuya función, no conocida al completo, se asocia con la regulación de la traducción del ARN a proteínas. Su función es vital para el correcto desarrollo cerebral, y prueba de ello es que las mutaciones de pérdida de función en FMRP causan el síndrome de X frágil (FXS), la forma más común de retraso mental congénita caracterizada, entre otras cosas, por comportamientos autistas y convulsiones (Verkerk y cols., 1991). De hecho, la implicación de los sitios de unión de esta proteína ya se había observado previamente en autismo (Iossifov y cols., 2012). Numerosos trabajos relacionan a FMRP con la plasticidad sináptica y la represión de la

traducción durante la elongación, lo que, sumado a las otras asociaciones con proteínas sinápticas (ARC y NMDAR), dibujan en mayor detalle el rol de la sinapsis en el desarrollo de la esquizofrenia. Varios estudios han descrito los genes a los que se une esta proteína a lo largo del genoma (Ascano y cols., 2012; Darnell y cols., 2011).

Tabla 1. 5. Mutaciones *de novo* en esquizofrenia, y comparación con autismo y retraso mental, en las tres categorías de genes con resultados más significativos. Los estudios comparados en esquizofrenia son los que aparecen en la tabla, y en autismo y retraso mental son los indicados en Fromer y cols. (2014). NS = variante no sinónima; LoF = Variante de pérdida de función; entre paréntesis se indica el número de genes afectados en cada categoría para cada estudio o grupo de estudios (Fromer y cols., 2014).

Set de genes			Sitios de unión de FMRP	Complejo ARC	Complejo NMDAR
Genes (N)			784	28	60
Fromer y cols. (2014)	NS (482)	Valor P	0.0094	0.00048	0.025
		N _{var}	64	6	6
	LoF (64)	Valor P	0.37	0.005	0.035
		N _{var}	7	2	2
Gulsuner y cols. (2013)	NS (68)	Valor P	0.065	1	1
		N _{var}	11	0	0
	LoF (12)	Valor P	1	1	1
		N _{var}	0	0	0
Xu y cols. (2012)	NS (137)	Valor P	0.027	1	0.13
		N _{var}	21	0	2
	LoF (20)	Valor P	0.55	1	0.086
		N _{var}	2	0	1
Estudios Autismo*	NS (789)	Valor P	0.003	0.22	0.031
		N _{var}	102	3	8
	LoF (134)	Valor P	0.0003	0.22	0.46
		N _{var}	26	1	1
Estudios Retraso mental*	NS (141)	Valor P	2.0 x 10⁻⁵	2.0 x 10⁻⁵	2.0 x 10⁻⁵
		N _{var}	40	5	8
	LoF (34)	Valor P	6.80 x 10⁻⁴	1.50 x 10⁻³	2.0 x 10⁻⁵
		N _{var}	10	2	5

Otro hecho destacable del estudio de Fromer y cols. (2014) es que las mutaciones *de novo* de pérdida de función, que de por sí no están sobrerrepresentadas en esquizofrenia, sí lo hacen cuando se seleccionan los pacientes con mayor fracaso escolar, arrojando la posibilidad de un impacto en el neurodesarrollo que explicaría la diversidad fenotípica asociada a ellas. En cualquier caso, parece claro que el riesgo en esquizofrenia es menor en comparación con autismo y, sobretudo, retraso mental (**tabla 1.5**). Este gradiente en la susceptibilidad de las

variantes raras en estos grupos de genes tiene sentido desde la perspectiva del trastorno de neurodesarrollo, donde la esquizofrenia representaría la versión leve y el retraso mental la más grave. Una evidencia adicional de esta hipótesis es el hecho de que las mutaciones de novo en esquizofrenia tienden a ocurrir en genes con mutaciones de novo reportadas en autismo y retraso mental (McCarthy y cols., 2014), ampliando la resolución del efecto pleiotrópico de estas variantes del nivel de locus (en estudios de CNV) al de gen.

El otro gran estudio de exoma en esquizofrenia (Purcell y cols., 2014), que se llevó a cabo mediante la secuenciación de 2536 casos y 2543 controles, es el mayor hasta la fecha. En él, la acumulación de variantes raras y funcionales no logró significación estadística en ningún gen tras las correcciones por múltiples test. Sin embargo, tras seleccionar un total de 2546 genes de categorías previamente implicadas en la enfermedad, se alcanzaron asociaciones significativas tanto para el total de genes, como para muchas de los sets seleccionados para el estudio. La acumulación de variantes de pérdida de función disruptivas era consistente en diferentes umbrales de frecuencia (**tabla 1.6**). Algo que, en menor medida, ocurría con las *NS strict* (variantes no sinónimas y patogénicas según cinco predictores de funcionalidad *in silico* diferentes). Las variantes *NS broad* (no sinónimas y patogénicas según al menos uno de los cinco predictores utilizados) ya no mostraron significación en el total de genes seleccionados, aunque sí lo hicieron para categorías concretas (**tabla 1.6**).

Del estudio de Purcell y cols. Se obtienen unos resultados que solapan en gran medida con los de Fromer y cols (2014), y otros anteriores (Girard y cols., 2011; Xu y cols., 2012a), para variantes *de novo*. De nuevo, las proteínas post sinápticas de los complejos glutamatérgicos ARC y NMDAR (aunque este último no pase la corrección por múltiples test) se ven implicados (Tabla 6), con valores de OR mayores que 5, así como los lugares de unión de la proteína FMRP. En este estudio, además, también se observa un solapamiento con uno de los más importantes y consistentes hallazgos de los estudios de GWAS, que lo constituyen los miembros de los canales de Calcio.

Este estudio también demostró, por primera vez en trabajos de secuenciación, la existencia de un componente poligénico para el riesgo asociado a estas variantes de baja frecuencia. Esta contribución de cientos de genes al desarrollo de la esquizofrenia es, sin embargo, un orden de magnitud menor al hallado en variantes comunes (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

Como conclusión a la revisión de los estudios de secuenciación en esquizofrenia, se pueden sacar dos conclusiones mayores. En primer lugar, los tamaños de muestras utilizados siguen resultando insuficientes para alcanzar la potencia estadística necesaria para variantes de riesgo menor. En segundo lugar, la perspectiva exómica de la esquizofrenia ha resultado

bastante pobre en resultados, y no explica la varianza genética no común que se le podía presuponer en un principio. Es evidente que hacen falta estudios de genoma completo u otros de resecuenciación de regiones específicas con potencial papel regulador, de forma que nos permita tener una visión de la variación rara más allá de las regiones codificantes.

Tabla 1. 6. Valores P referidos a la acumulación de variantes en los 2546 genes seleccionados por su implicación previa en esquizofrenia, así como en cada una de las categorías por separado. En negrita se resaltan los valores significativos tras correcciones por múltiples test. Datos extraídos de Purcell y cols. (2014).

Tipo de variante	Set	N genes	singletons	MAF < 0.1%	MAF < 0.5%
<i>Disruptive</i>	Toda la selección	2546	0.0008	0.0001	0.0002
<i>NS_{strict}</i>			0.0059	0.0015	0.011
<i>NS_{broad}</i>			0.0986	0.1295	0.1126
<i>Disruptive</i>	SCZ de novo				
	Seq Exoma (<i>disruptive</i>)	87	0.0319	0.0007	0.0003
	Seq Exoma (NS)	611	0.0053	0.0011	0.0055
	CNV				
	genes CNV <i>de novo</i>	234	0.0234	0.0039	0.0124
	genes CNV en SCZ	345	0.3308	0.4596	0.4376
	GWAS				
	Canales de Calcio	26	0.0019	0.0214	0.0212
	SNP ($P < 10^{-4}$)	479	0.1794	0.0368	0.0037
	sitios de unión de MIR137	446	0.6573	0.5609	0.4747
	Sinapsis				
	Densidad post sináptica	685	0.0808	0.1154	0.1256
	ARC	28	0.0016	0.0014	0.0014
	NMDAR	61	0.0158	0.0251	0.0252
	PSD-95	65	0.0017	0.0009	0.001
	mGluR5	39	0.1327	0.09	0.0902

1.2.3. Arquitectura genética de la esquizofrenia

El desarrollo tecnológico y las consecuentes técnicas desarrolladas durante los últimos años para el estudio del componente genético de la esquizofrenia no han conseguido, pese a los grandes descubrimientos al respecto, encontrar respuestas para la elevada heredabilidad que se le atribuye. La complejidad de la esquizofrenia en el campo clínico y del diagnóstico se ha demostrado igual o mayor desde el punto de vista genético, lo que puede explicar la sintomatología compartida y comorbilidad elevada con otras patologías psiquiátricas (Craddock y Owen, 2010; Doherty y Owen, 2014). Sin embargo, a día de hoy, podemos sacar algunas conclusiones importantes acerca de la arquitectura genética de la enfermedad que, en última

instancia, pueden ser útiles a la hora de determinar la dirección que se debería elegir para futuros esfuerzos en el estudio de esta enfermedad, y conseguir implementar un diagnóstico clínico y unas políticas de prevención basadas, en una mayor medida, en aspectos genéticos concretos.

1.2.3.1. Hipótesis variante común – enfermedad común Vs hipótesis variante rara – enfermedad común

Los estudios familiares, en gemelos y de adopción han servido para demostrar una muy alta heredabilidad en la esquizofrenia (65 - 80%), pero la falta un patrón de herencia determinado y el descubrimiento progresivo de muchos *loci* relacionados con la enfermedad revelaron que la esquizofrenia es una patología multifactorial compleja, al igual que otros trastornos psiquiátricos (O'Donovan y cols., 2009).

Esta complejidad desde el punto de vista genético fue abordada, teóricamente, desde dos hipótesis diferentes, mediante las cuales se buscaba explicar la alta incidencia en la sociedad (Messias y cols., 2007). Una defiende la susceptibilidad genética provocada por variantes comunes de frecuencia elevada en la población y otra atribuye el riesgo principalmente a variantes raras de origen reciente.

La **hipótesis enfermedad común- variante común** surgió en primer lugar (Lander y cols., 1996). Como consecuencia de la elevada prevalencia de la esquizofrenia, diversos investigadores pensaron que la patología debía estar causada principalmente por muchas variantes de bajo riesgo que alcanzaban una frecuencia elevada en la población (> 1%) al no ser, de forma individual, factores de riesgo suficientemente elevados como para estar sometidas a selección purificadora.

Los estudios de asociación son una estrategia más potente para descubrir variantes genéticas de susceptibilidad a esquizofrenia que obedezcan este contexto teórico (Risch y Merikangas., 1996). El PGH permitió conocer a completo la secuencia del genoma. Ello sumado al desarrollo de *arrays* de genotipado, cada vez con mayor resolución, y del proyecto HapMap, que supuso el conocimiento de la estructura haplotípica a lo largo del genoma, permitió la realización de los estudios de GWAS (mediante asociación indirecta por DL), alcanzando tamaños de muestras muy elevados en los últimos estudios (Ripke y cols., 2013a; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

Los estudios de asociación revelaron la implicación de más de 100 *loci* independientes, que englobaban 350 genes y varias regiones intergénicas, lo cual demostró que las bases

argumentales de esta teoría acerca de la existencia de variación genética común eran ciertas. Sin embargo, y pese a los más de 50000 individuos esquizofrénicos utilizados en el GWAS de mayor tamaño (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), el componente poligénico no supera el 7%, lo que se aleja notablemente de las estimaciones de heredabilidad. Por tanto, a día de hoy existen suficientes indicios para pensar que la hipótesis enfermedad común – variante común podría no ser la única explicación de la susceptibilidad genética subyacente de la esquizofrenia.

Por otra parte, otros autores defendieron la importancia de la variación rara o infrecuente como mecanismo genético mayoritario de riesgo en la esquizofrenia (Manolio y cols., 2009; McClellan y cols., 2007; Pritchard, 2001). Bajo la **hipótesis enfermedad común - variante rara (ECVR)**, múltiples variantes raras de susceptibilidad de riesgo moderado estarían influyendo en el desarrollo de la esquizofrenia, sin que ninguna de ellas fuera por sí sola determinante para el fenotipo. La esquizofrenia es una enfermedad que aparece predominantemente durante la adolescencia e inicio de la adultez, lo que es consistente con la fertilidad reducida asociada a ella (Haukka y cols., 2003). Esto genera una selección negativa de este fenotipo que contrastaría con la persistencia de variantes comunes a lo largo de muchas generaciones, pues la selección purificadora no permitiría que estas alcanzasen frecuencias elevadas si realmente están asociadas con el riesgo a padecer la enfermedad. En este sentido, estos autores plantean la existencia de variantes *de novo* en la línea germinal (Crow, 2003) o de aparición reciente en la población que desaparecerían a lo largo de pocas generaciones, o bien segregarían en la población a frecuencias bajas cuando su riesgo asociado fuese menor (Gorlov y cols., 2003).

El desarrollo y aumento de resolución de las tecnologías de genotipado han permitido la búsqueda de variantes estructurales de tamaño menor al detectable por las técnicas de citogenética clásica, lo que demostró la gran contribución de las CNV a la susceptibilidad genética de la esquizofrenia (Malhotra y cols., 2012; Rees y cols., 2014; Walsh y cols., 2008). Estos hallazgos revelaron la predisposición genética de variantes raras de riesgo elevado a este y otros trastornos psiquiátricos, así como a trastornos severos del desarrollo, siendo una demostración tanto de la contribución de la variación rara al trastorno psiquiátrico como de la naturaleza de la esquizofrenia como trastorno del neurodesarrollo (Walsh y cols., 2008).

Por su parte, la contribución de la variación rara puntual, a nivel de nucleótido, pudo estudiarse con el reciente impulso al desarrollo de las plataformas de secuenciación de NGS, lo que ya ha demostrado el papel de lo que podrían ser cientos o miles de *loci* a lo largo de miles de exomas de pacientes que, albergando variación rara, se encuentran implicados en la enfermedad (Fromer y cols., 2014; Purcell y cols., 2014). La demostración de la existencia de

un componente poligénico de un orden de magnitud inferior al encontrado en la variación común (Purcell y cols., 2014) pone de manifiesto la multifactorialidad genética en variación rara así como la pobre explicación de la heredabilidad conseguida con estos estudios.

Desde el punto de vista de la arquitectura genética, se puede concluir que hay suficientes razones para afirmar que ambas hipótesis no son excluyentes (Goldstein y Chikhi, 2002), sino que el riesgo causado por la aparición de una variante rara de efecto medio podría ser o no determinante para el fenotipo en función de la combinación con la variación común en genes de susceptibilidad a esquizofrenia, así como de las interacciones epistáticas entre los genes afectados.

En la **figura 1.7** se representa la dificultad subyacente de los estudios genéticos para la detección de variantes asociadas en función de la naturaleza de las mismas. Variantes de alto riesgo y comunes en la población son muy fáciles de detectar pero escasas (y no existentes en esquizofrenia), mientras que variantes raras de efecto muy bajo requieren de enormes tamaños de muestra en estudios de asociación para su detección, por lo que tampoco interesa su búsqueda dada su escasa repercusión en el fenotipo. Las dos líneas de puntos encierran la varianza genética más relevante, que va desde variantes comunes de bajo riesgo (hipótesis ECVC, estudiable mediante estudios de asociación) hasta variantes raras de alta predisposición (hipótesis ECVR, con estudios de CNV, ligamiento y secuenciación), pasando por variantes de frecuencia y riesgo intermedios (estrategias específicas para cada caso).

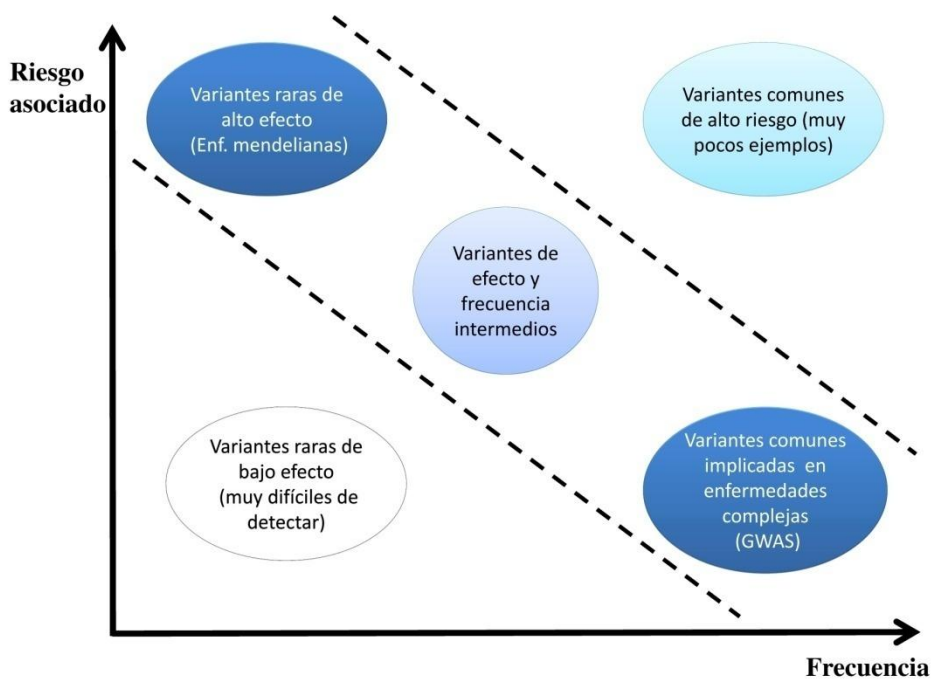
1.2.3.2. Heredabilidad perdida

La conclusión más importante, sin embargo, que se puede extraer de los múltiples estudios genéticos en esquizofrenia es que, pese al gran número de resultados satisfactorios, no se ha conseguido explicar la mayor parte de la heredabilidad que atribuye a la varianza genética entre un 65 y un 80% de la varianza fenotípica.

Una de las posibles causas valoradas es la sobreestimación de la heredabilidad en los cálculos realizados. Esto podría ser debido, por ejemplo, a no haber conseguido separar completamente el ambiente compartido en los estudios de adopción. Sin embargo, pese a que algunas estimaciones pudieran estar infladas, la mayor parte de la comunidad científica es partidaria de la idea de que esto no sería suficiente como para reducir el enorme margen de diferencia entre la heredabilidad explicada y la estimada (Maher, 2008). Por otra parte, no hay que olvidar que la heredabilidad es un concepto propio de cada población, de modo que una población sometida a un factor ambiental de riesgo medio para el desarrollo de la esquizofrenia, tendría una heredabilidad menor que otra población en donde la contribución del ambiente fuera

prácticamente nula. Por esa razón, tampoco sería descabellado pensar que muchos estudios realizados en países con un ambiente más "adverso", podrían conducir a infraestimaciones, si de ella se extrapola el valor de la heredabilidad media interpoblacional.

Figura 1. 7. Viabilidad en la detección de variantes genéticas en función del riesgo y de la frecuencia de las mismas en la población. Figura adaptada de Manolio y cols. (2009).



Otro factor que podría estar detrás de las heredabilidades perdidas en trastornos complejos es la epigenética (Trerotola y cols., 2015). Pese a que las marcas epigenéticas como la metilación del genoma y las modificaciones químicas de las histonas se eliminan en su mayoría tras la fecundación y otros momentos específicos como la meiosis gametogénica, se ha visto que algunos patrones epigenéticos permanecen y podrían heredarse del mismo modo que los cambios en la secuencia nucleotídica (Gkoutela y cols., 2015; Smith y cols., 2014), incluyéndose por tanto en la varianza genética (Trerotola y cols., 2015). No obstante, la aportación puramente epigenética parece bastante discutible. A parte de la dificultad de escindir la herencia epigenética de modificaciones en el epigenoma no heredadas, sino como medio biológico por el cual las células ejecutan su adaptación rápida a los cambios ambientales, diversos estudios han revelado que las diferencias en las modificaciones histónicas u otras modificaciones químicas podrían tener su origen en variantes genéticas situadas en regiones reguladoras (Kilpinen y cols., 2013; McVicker y cols., 2013). En este sentido, los análisis más recientes de los resultados de GWAS demuestran que los *loci* asociados a enfermedades

psiquiátricas están enriquecidos en genes implicados en la modificación química de histonas (Network, T., & Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015).

Por ello, si bien los estudios de asociación consiguen capturar la variación común a lo largo de todo el genoma, los estudios de secuenciación se han focalizado en la variación codificante como primera aproximación, lo que deja una pieza fundamental del puzzle genético que no ha sido tomada en cuenta. Del mismo modo que los estudios de GWAS demuestran la contribución de mecanismos de regulación fuera del exoma, la variación rara en regiones reguladoras, mediante resecuenciación de regiones específicas o del genoma completo, debe ser el siguiente paso a dar en la búsqueda de las variantes de susceptibilidad a esquizofrenia. Tal vez, la respuesta a la heredabilidad no explicada todavía se encuentre en la contribución de la hipótesis enfermedad común – variante rara más allá de la secuencia codificante.

1.3. REGULACIÓN DE LA EXPRESIÓN GÉNICA EN ESQUIZOFRENIA

1.3.1. Estrategias para abordar un estudio de regulación de la expresión génica en esquizofrenia

Con las numerosas evidencias de que una gran parte de la genética de la esquizofrenia se debe a variabilidad en regiones no codificantes, se hace insuficiente el valor de la secuencia del genoma humano y más evidente la necesidad de un conocimiento mucho más exhaustivo acerca de los elementos funcionales que éste alberga. Igualmente, se requiere información acerca de cómo interactúan entre ellos para dar lugar a los diferentes fenotipos característicos de las enfermedades complejas como en el caso de la esquizofrenia u otros trastornos psiquiátricos.

De los más de 10000 SNP asociados con diferentes fenotipos en los estudios de GWAS hasta la fecha, hasta un 88% se encuentran fuera de las regiones codificantes (Hindorff y cols., 2009). Sin embargo, la mayoría no ofrece ninguna evidencia biológica de cómo puede afectar al desarrollo de la enfermedad. La variación común de muchas enfermedades complejas está fuertemente relacionada con la regulación genética (Maurano y cols., 2012), pero así como diversos algoritmos de predicción funcional permiten el análisis *in silico* de variantes codificantes no sinónimas, la interpretación de la posible repercusión de variantes reguladoras es más complicada.

En el caso de la esquizofrenia, solo 10 de las más de 100 regiones asociadas se pueden atribuir de forma fidedigna a cambios en la secuencia aminoacídica (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), lo cual ya nos indica la máxima prioridad que tiene el conocer la funcionalidad de las regiones no codificantes del genoma para

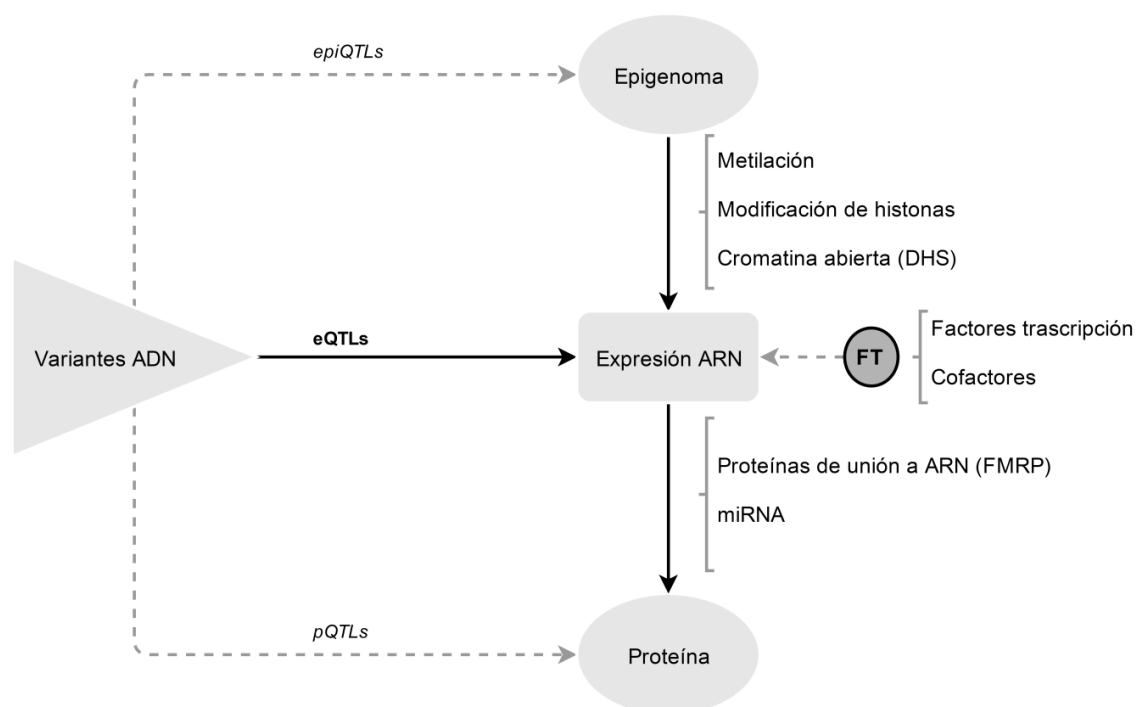
inferir la posible causa biológica de las enfermedades con las que se asocian. Del mismo modo, otros elementos como los genes diana de la proteína FMRP (Fromer y cols., 2014; Purcell y cols., 2014) o los que regulan las modificaciones químicas en las histonas (Network, T., & Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015), demostraron que la regulación post transcripcional y de los mecanismos epigenéticos, respectivamente, están muy relacionados con el desarrollo de la esquizofrenia. Es por ello que la heredabilidad perdida característica de la esquizofrenia podría deberse a la variación genética rara en regiones reguladoras, algo que no se ha podido capturar con los métodos implementados hasta la fecha entre otras cosas por la complicada interpretación y escasa información que podrían aportar las posibles asociaciones. De cara a conocer la arquitectura de la regulación de expresión en la esquizofrenia, es necesario abordar el estudio desde distintos puntos de vista (**figura 1.8**).

- En primer lugar, la expresión génica ocupa una posición clave en el flujo de la información biológica de un organismo. Por un lado, representa el *output* de los mecanismos genéticos y epigenéticos de regulación que operan a nivel de la secuencia de ADN y a través de la interacción con el entorno. Por otro lado, representa el *imput* de la mayoría de procesos biológicos que se valen de la presencia de proteínas y otras especies moleculares para tener lugar. Si bien la utilización de organismos modelo es una opción muy adecuada para el estudio de vías moleculares determinadas, en el caso del sistema nervioso central (SNC) y particularmente el cerebro, es más complicado extrapolar resultados de los mecanismos de regulación de estudios animales (Johnson y cols., 2009; Nord y cols., 2015), pudiendo no corresponderse los fenotipos con mutaciones similares (Liao y Zhang, 2008). Es por ello que el conocimiento del transcriptoma del cerebro humano en el contexto de la esquizofrenia es esencial para entender qué funciones pueden verse alteradas a raíz de la expresión diferencial de determinados genes.
- En segundo lugar, dada la importancia de algunos genes en el contexto del neurodesarrollo cerebral, es muy factible que la arquitectura del proceso de transcripción de estos genes se haya conservado a lo largo de la evolución o haya sufrido cambios en el linaje humano. El rastreo de genes o regiones próximas a ellos, cuya secuencia ha sido conservada a lo largo de la evolución de los vertebrados, mamíferos o primates, puede permitir seleccionar *loci* de elevado potencial regulador para un análisis más profundo de la variación genética en el contexto de las enfermedades complejas.
- Finalmente, Dada la implicación de regiones no codificantes en la genética de la esquizofrenia, es preciso conocer la estructura funcional del genoma a nivel de regulación de la expresión. La transcripción es un proceso muy complejo en el que

intervienen muchos elementos reguladores y la alteración de alguna de las etapas del mismo puede provocar cambios significativos en el fenotipo final. Por ello, es muy importante conocer los elementos funcionales del genoma para poder interpretar la variación genética presente en estos *loci*.

La integración de los estudios genéticos y de expresión permite localizar las regiones del genoma que albergan variantes que influyen en los niveles de expresión, denominadas eQTLs (del inglés *expression quantitative trait loci*), lo que ha sido objeto de muchos estudios en diversas enfermedades complejas durante los últimos años (Albert y Kruglyak, 2015). Esta influencia puede deberse a variación genética que actúa directamente o en *cis*, como en la alteración de la secuencia de unión de un factor de transcripción en el promotor del gen que regula, o de forma indirecta o en *trans*, como por ejemplo mediante la variación en la secuencia de un cofactor cuya estructura alterada reducirá la eficiencia de la transcripción de los genes que regula. Del igual modo, la variación también puede ser reguladora por alterar la dinámica de los estados de la cromatina u otros patrones epigenéticos (Johannes y cols., 2008), dando lugar a los *epigenetic quantitative trait loci* (*epiQTLs*) o por influir en los niveles de proteína (Melzer y cols., 2008), caso de los *protein quantitative trait loci* (*pQTLs*). En cualquier caso, la caracterización de los elementos funcionales del genoma es condición necesaria para interpretar la variación genética en regiones no codificantes del genoma (**figura 1.8**).

Figura 1. 8. Posibles vías de influencia de la variación genética de regiones reguladoras en la expresión.



1.3.2. Estudios de expresión en esquizofrenia

1.3.2.1. Técnicas en el estudio de la expresión génica.

Los primeros estudios para comparar la expresión génica en cerebros de esquizofrénicos y controles sanos se realizaron en **microarrays de expresión**, similares a los de genotipado, pero en los que se cuantifica la unión de los genes transcritos mediante sondas complementarias a la secuencia de los mismos. De este modo, los niveles de expresión de miles de genes se controlan simultáneamente para estudiar los efectos de ciertos tratamientos, enfermedades y/o etapas de desarrollo sobre la expresión génica. La tecnología de *microarrays* también se puede utilizar para medir la expresión en todo el transcriptoma y así identificar genes con niveles de expresión significativamente diferentes en cerebros de casos y controles, de modo que se puedan generar hipótesis sobre el posible riesgo de determinados genes para el desarrollo de la enfermedad. El continuo desarrollo de esta tecnología ha permitido cuantificar isoformas de *splicing* alternativo mediante sondas que abarcan la unión de diferentes exones (Clark y cols., 2002).

Sin embargo, son muchas las limitaciones de los *microarrays* de expresión. Pese a su aceptable coste, en chips de alta resolución, el coste económico se dispara y se hace inviable para muchas aplicaciones. En segundo lugar, la alta dependencia del conocimiento de la secuencia de los genes analizados origina problemas cuando hay variaciones en ella en determinados tejidos o líneas celulares no conocidas. Por otra parte, dado que se fundamenta en la unión de la sonda con la secuencia de interés, es muy habitual el ruido de fondo producto de la hibridación inespecífica (Okoniewski y Miller, 2006). Esto reduce la sensibilidad para la medición de genes de baja expresión, y en conjunción con la saturación fotométrica a elevados niveles, reduce el rango dinámico en el cual se pueden cuantificar las distintas especies de ARN.

Otros métodos se valieron de la secuenciación de Sanger para el estudio de la expresión génica con el **análisis de marcadores de secuencia expresada o EST** (del inglés *Expressed sequence tags*), mediante los cuales subfragmentos de los genes eran secuenciados y cuantificados para inferir de ellos la expresión de los genes. En este sentido, técnicas como SAGE, del inglés *Serial analysis of gene expression* (Velculescu y cols., 1995), o CAGE (Shiraki y cols., 2003), del inglés *Cap based analysis of gene expression*, mediante la cual se secuencian el extremo 5' de los ARNm con proteína CAP en su extremo, son algunos ejemplos. Estos métodos demostraron un alto rendimiento y mayor sensibilidad y precisión que los *microarrays* de expresión, pero las limitaciones de la secuenciación de Sanger y la inespecificidad de isoformas que comparten las mismas *tags*, son factores que limitaron su uso universal.

La aplicación de las tecnologías de ultrasecuenciación (NGS) a los estudios de expresión supuso una auténtica revolución por la combinación de su alto rendimiento y reducido coste (Wang y cols., 2009b). La **ultrasecuenciación de ARN** o **RNA-seq** tiene lugar mediante la secuenciación de las especies de ARN fragmentadas, previa conversión a ADN complementario, con las plataformas de NGS utilizadas en la secuenciación de ADN (**tabla 1.4**). Conlleva, además, una serie de importantes ventajas respecto a los métodos predecesores.

- En primer lugar, en lugar de la detección fotométrica de los chips de expresión, la cuantificación de los genes se hace mediante el cálculo de la profundidad de cobertura en el mapeo frente a genoma o transcriptoma de referencia. Este mapeo frente a una secuencia de referencia es mucho más específico que la hibridación a las sondas de los *microarrays*, y por ello el ruido de fondo es mucho menor, pudiendo analizar genes cuyos niveles de expresión son reducidos. Del mismo modo, esta metodología no presenta un límite máximo de detección en tanto que no ocurre la saturación propia de detectores fotométricos. Por estas características, la RNA-seq presenta un amplio rango dinámico en el que cuantificar diferencias de expresión entre diferentes genes.
- Como alternativa al mapeo frente a un genoma de referencia, es posible usar las lecturas para el ensamblaje de un transcriptoma *de novo*, lo cual es muy interesante para descubrir especies o isoformas nuevas presentes en la transcripción de determinados tejidos. Así, el análisis del transcriptoma no se limita a las especies cuya secuencia es conocida.
- La resolución del método es máxima (a nivel de base nucleotídica), por lo que constituye la mejor técnica para el análisis de la conectividad entre exones y de isoformas de splicing alternativo.
- Se trata de una técnica muy precisa, tal y como ha sido comprobado mediante RT-PCR (Nagalakshmi y cols., 2008), altamente reproducible (Cloonan y cols., 2008), lo que permite la comparación de resultados de diferentes estudios con mayor facilidad que con otros métodos, y con baja utilización de muestra de ARN.
- Finalmente, y como algo muy innovador respecto a las técnicas homólogas, permite el estudio de la variación genética a escala de transcriptoma (Piskol y Ramaswami, 2013). A pesar de no cubrir la totalidad del genoma, permite un análisis de los procesos de modificación de la secuencia de ARN post transcripcional o *RNA editing* (Ramaswami y cols., 2013) y la comparación con estudios de WGS o WES sirve como validación de los mismos. En el caso de la esquizofrenia y otras enfermedades psiquiátricas como el autismo, es necesario señalar que uno de los grupos de genes asociados por la presencia de variantes raras son las dianas de las proteínas de unión a ARN como FMRP (De Rubeis y cols., 2014; Purcell y cols., 2014) o FBOX1 (De Rubeis y cols., 2014),

frecuentemente unidas a las regiones codificantes, y con efectos sobre la regulación de la traducción. Por otra parte, la variación genética en los lugares de unión de miRNAs implicados en esquizofrenia (Perkins y cols., 2007), así como en algunas especies de miRNAs y en particular MIR137 (Ripke y cols., 2013a), se pueden estudiar mediante RNA-seq permitiendo, además, comprobar si las especies mutadas se expresan en el tejido de estudio y tienen mayor evidencia de ser patogénicas. Un estudio reciente demostró la presencia de sitios de unión de miRNAs en cerebro humano en más de 3000 genes distintos (Boudreau y cols., 2014), constituyendo una herramienta muy valiosa para el estudio combinado de la variación genética que afecte a esos lugares de unión y la expresión de los genes diana, analizando el efecto *in situ* de esas variantes.

De igual modo que con la aplicación de las tecnologías de NGS en el estudio de ADN, los resultados provenientes de RNA-seq requieren de software sofisticado por el volumen de información y la necesidad de alcanzar unos estándares de calidad aceptables. En los análisis de expresión entre muestras de casos y controles, los archivos de secuencias mapeadas frente al genoma o transcriptoma de referencia se analizan mediante diversas pruebas estadísticas para determinar la presencia de genes diferentemente expresados (DE) entre casos y controles. Existen muchos *software* con algoritmos más o menos parecidos en los que se puede implementar este análisis, pero se puede dividir en dos tipos principales:

- Métodos que analizan la **expresión a nivel de gen**: Mediante estos métodos se cuantifican las especies deseadas (por ejemplo, todos los genes de RefSeq o Gencode), indicando las posiciones que ocupan en el genoma. Estos programas ejecutan la comparación de expresión mediante el análisis de la distribución binomial de las lecturas de los genes en casos y controles, por lo que se conocen como métodos basados en lecturas (*count-based methods*). *EdgeR* (Robinson y cols., 2010) y *DESeq* (Anders y Huber, 2010) son los principales software más utilizados en este sentido.
- Métodos que analizan la **expresión a nivel de transcrito**: Estos software de diseño más reciente permiten la cuantificación a nivel de isoforma transcrita, lo que presenta la principal ventaja de conocer la arquitectura del splicing en un tejido concreto, y su posible relación con la condición o enfermedad estudiada. *Cufflinks* (Trapnell y cols., 2012) es el principal ejemplo de este tipo. A diferencia de los anteriores, utiliza como medida de expresión fragmentos por kilobase de longitud y millón de lecturas mapeadas (FPKM), lo que permite una normalización de los distintos grados de expresión de los transcritos y de la longitud de los mismos, y posibilita comparar la expresión de genes diferentes.

Un estudio reciente, en el que se analizaron los parámetros de rendimiento y calidad de los tres software citados, demostró que, si bien *Cufflinks* presenta una clara ventaja sobre la utilidad en el estudio de nuevas isoformas, las metodologías basada en lecturas de *EdgeR* y *DESeq* mostraron una mayor precisión cuando se estudiaba la expresión a nivel de gen (Zhang y cols., 2014b). Esto indica que ambas posibilidades son complementarias, y la elección de una u otra dependerá del tipo de estudio deseado.

Por otra parte, en los últimos años han sido varios los proyectos que se han valido de los avances técnicos y los mayores tamaños muestrales para tratar de describir el transcriptoma a través de las dimensiones espaciales y temporales, de forma que se pueda mejorar el conocimiento al respecto mediante una visión global de los patrones de expresión génica en los diferentes tejidos y a lo largo de las distintas etapas vitales. Gracias a estos esfuerzos, estas bases de datos de libre acceso constituyen herramientas muy buenas para tratar de interpretar resultados genéticos en combinación con datos de expresión.

Un ejemplo importante es el del proyecto GTEx (Lonsdale y cols., 2013). Este proyecto nació del impulso del Instituto nacional de salud estadounidense (NIH) con el propósito de conocer la expresión genética a lo largo de más de 50 tejidos del cuerpo humano en cientos de individuos y, además, mediante el análisis combinado con herramientas de genotipado, proporcionar información de la variabilidad genética común entre individuos para detectar y caracterizar los eQTLs que influyen en la expresión génica en cada tejido (Ardlie y cols., 2015). En el contexto cerebral, destaca también el *BrainSpan Atlas Atlas* (www.brainspan.org), un proyecto que ha conseguido desvelar los patrones de transcripción cerebral de los genes a lo largo del desarrollo, algo que ha servido para demostrar que el 84% de los genes se transcriben en el cerebro en algún momento de la vida (Hawrylycz y cols., 2012).

Otros estudios han volcado sus esfuerzos en desvelar qué genes son específicos de los tipos celulares cerebrales (Cahoy y cols., 2008; Zhang y cols., 2014a) y cuáles se coexpresan siguiendo patrones similares a lo largo de la vida humana (Hawrylycz y cols., 2012) o durante neurodesarrollo fetal (Miller y cols., 2014). Con los resultados de estos trabajos, se ha demostrado que muchos genes que se coexpresan siguiendo un perfil determinado son más o menos específicos de según qué tipo celular, estableciéndose relaciones muy reveladoras entre el momento de la transcripción y el tipo celular implicado (Mistry y cols., 2013b). Xu y cols., sirviéndose de los datos de RNA-seq de *BrainSpan*, diseñaron una novedosa herramienta denominada CSEA (del inglés *Cell - type enrichment analysis*), mediante la cual se puede analizar si un conjunto de genes de interés está significativamente enriquecido en aquellos específicos de tipos determinados celulares en distintas áreas cerebrales y en distintos momentos del desarrollo (Xu y cols., 2014). Mediante este método, los autores demostraron que

los genes sobreexpresados en muestras de cerebros de individuos con autismo son característicos de astrocitos, oligodendrocitos y células del sistema inmune, mientras que los genes candidatos en autismo por estudios genéticos previos eran característicos de neuronas corticales y talámicas.

La utilización de estas bases de datos en estudios genéticos ha servido para revelar la conexión entre patrones de expresión y variantes patogénicas en esquizofrenia (Gilman y cols., 2012; Xu y cols., 2012a). Tanto los trabajos de Xu y cols. como los de Gilman y cols. demostraron una acumulación significativa de las variantes genéticas en genes altamente expresados durante el desarrollo prenatal. Igualmente, otro estudio coetáneo demostró que módulos de coexpresión enriquecidos en genes neuronales (de sistemas de GABA y glutamato) y de oligodendrocitos presentaban una acumulación estadísticamente significativa de variantes genéticas de riesgo a esquizofrenia (Roussos y cols., 2012). Posteriormente, Gulsuner y cols., mediante el estudio combinado del exoma de más de 100 tríos con datos de expresión espaciotemporal de *BrainSpan*, también demostraron que las variantes *de novo* patogénicas se acumulaban significativamente en genes críticos de la neurogénesis en córtex prefrontal (Gulsuner y cols., 2013).

1.3.2.2. Genes diferentemente expresados (DE) en esquizofrenia

En el contexto de la esquizofrenia, muchos estudios se han llevado a cabo para determinar la expresión diferencial en casos y en controles a partir de muestras de cerebro *post mortem*.

Los primeros casos de trabajos realizados para buscar diferencias de expresión en *microarrays* fueron escasamente replicados, habiendo importantes limitaciones como la diferencia entre plataformas utilizadas, la edad de la muestra empleada, causa de muerte o protocolos estadísticos diferentes (Sequeira y cols., 2012). Estas limitaciones, que eran más propias de los métodos inicialmente empleados, también son extensibles en menor medida a los actuales, aunque las más importantes son la calidad y el tamaño de muestra (Mistry y cols., 2013a). La dificultad para obtener tejido cerebral humano *post-mortem* de alta calidad y la ausencia de tamaños de muestra suficientes para superar la variabilidad individual son los grandes inconvenientes de la aplicación de tecnologías genómicas en estos estudios. Es por ello que los estudios de expresión comparada caso - control no han utilizado tamaños de muestras mayores a treinta individuos, llegando en casos específicos de colaboraciones a poco más de cien (Mistry y cols., 2013a).

A pesar de las limitaciones al respecto de los chips de expresión, muchos genes DE en pacientes esquizofrénicos fueron reportados en numerosos estudios. Los primeros trabajos estuvieron enfocados al estudio del córtex prefrontal, en concreto el área dorsolateral (DFC). Esta región, que es la última en mielinizarse del cerebro, fue objeto de estudio inicialmente en esquizofrenia debido a su continuo y activo desarrollo hasta la adolescencia y por su papel en la regulación de las funciones ejecutivas y de socialización del cerebro (Lenroot y Giedd, 2006; Orellana y Slachevsky, 2013), elementos clave en el fenotipo esquizofrénico. No obstante, pese a que se hallaron numerosos genes DE en DFC de enfermos (Choi y cols., 2008; Maycox y cols., 2009; Mirnics y cols., 2000), también los hubo de forma generalizada en otras áreas como el giro temporal superior (Bowden y cols., 2008), la corteza temporal (Aston y cols., 2004), el hipocampo (Altar y cols., 2005), la amígdala (Weidenhofer y cols., 2006) o células gliales (Catts y cols., 2014; Katsel y cols., 2005); por lo que la patofisiología cerebral característica de la esquizofrenia es extensible a prácticamente la totalidad del cerebro y, probablemente, la interconexión de las distintas áreas juegue un papel más importante que algunas en particular (Lynall y cols., 2010).

Diversas revisiones de estos estudios revelan que el conjunto de genes DE en esquizofrenia, al analizarse de forma integrada mediante análisis de vías o *pathways*, están enriquecidos en múltiples funciones biológicas entre las que destacan la respuesta inmune e inflamatoria, plasticidad sináptica, regulación de la transcripción y traducción, mielinización, función mitocondrial, función de la cromatina, neurogénesis y sistema de neurotransmisión del glutamato y de GABA (Kumarasinghe y cols., 2012; Sequeira y cols., 2012). Es interesante señalar el solapamiento existente entre los procesos desregulados en el transcriptoma cerebral y los hallazgos genéticos en genotipado y secuenciación, reforzando la idea de que, pese a la contribución ambiental y la variabilidad interpersonal, la mayoría de estos resultados podría tener una justificación principalmente genética.

Con la aparición de la tecnología de RNA-seq se replicaron muchos de los hallazgos previos en *microarrays* de expresión, siendo particularmente frecuentes los resultados atribuibles a genes de respuesta inflamatoria (Fillman y cols., 2013; Hwang y cols., 2013; Xu y cols., 2012). Es necesario señalar, en cualquier caso, que la credibilidad de estos hallazgos, pese a su relación con otros resultados genéticos y su alta reproducibilidad, se incrementará cuando los tamaños de muestras empleadas sean lo suficientemente grandes como para soslayar el problema de la variabilidad entre ellas. En este sentido, frecuentemente se han encontrado resultados contradictorios en muchos tipos celulares o procesos biológicos estudiados. Una de las interferencias a tener en cuenta es la de la medicación. Es casi imposible acceder a muestras de pacientes que no estén o hayan estado bajo el efecto de antipsicóticos, además de la comorbilidad con el abuso de otras sustancias como alcohol o cannabis que podrían alterar la

expresión de diversos genes. El uso de antipsicóticos afecta a la presencia de metabolitos y proteínas en cerebros *post mortem* (Chan y cols., 2011) y, además, también se ha demostrado que altera la expresión en cerebros de primates (O'Connor y cols., 2007). Por tanto, el consumo de estos medicamentos son una innegable fuente de falsos positivos y negativos que se deberán tener en cuenta a la hora de interpretar los resultados de expresión con cautela.

Diversos estudios demostraron que los genes DE en casos de esquizofrenia respecto a controles se encontraban enriquecidos en genes candidatos de la enfermedad y aquellos cuya expresión era dependiente de la edad (Choi y cols., 2009a; Harris y cols., 2009). Ejemplo de ello son algunos genes, como diversos receptores de glutamato, que alcanzan su mayor nivel de expresión en las etapas del neurodesarrollo u otros como *NRG1*, que presentan un pico de expresión en edades comprendidas entre los 15 y 25 años, englobando las edades de inicio de la mayoría de los casos. Por esta razón, es interesante conocer los perfiles de expresión de los distintos genes que se expresan en el cerebro para interpretar el efecto de las variantes genéticas en la transcripción. Un estudio posterior consolidó la credibilidad de estos indicios al revelar que los genes cuya expresión se reduce con la edad, propios de procesos del neurodesarrollo y la diferenciación, no mostraban ese perfil en casos de esquizofrenia (Torkamani y cols., 2010).

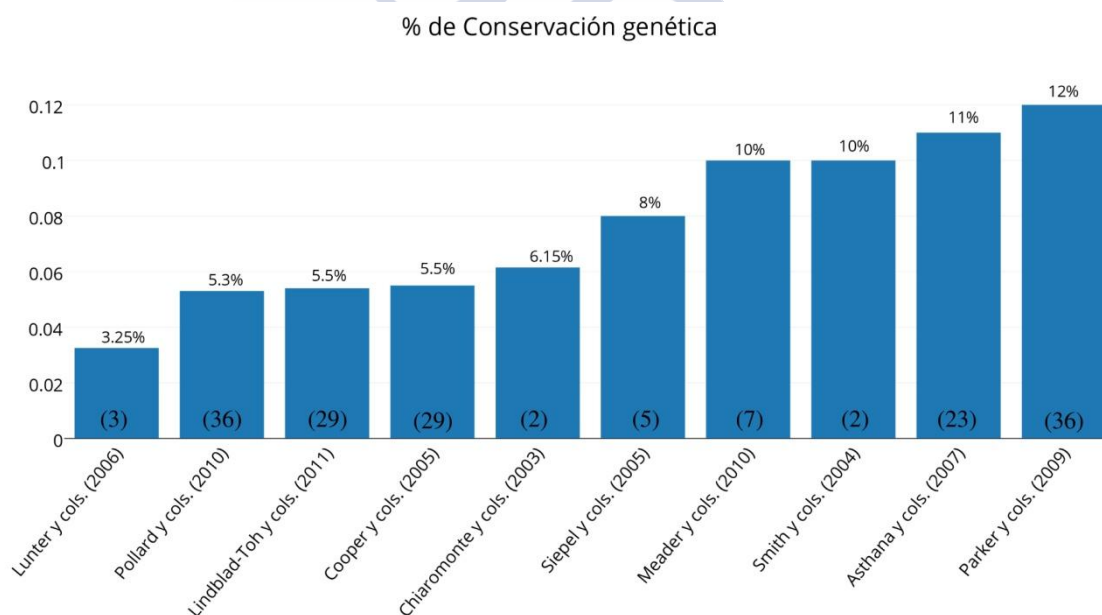
1.3.3. La importancia de la conservación genética en la identificación de elementos reguladores de la expresión génica.

1.3.3.1. Conservación genética en mamíferos

La conservación en la secuencia de nucleótidos a lo largo de la evolución humana ha sido considerada, clásicamente, como una de las aproximaciones más valiosas a la hora de buscar regiones funcionales en el genoma humano. Si un locus determinado presenta una función de vital importancia para la especie, cualquier mutación que afecte a su secuencia se eliminará en la población, estando la región sometida a una selección purificadora a lo largo de millones de años. Por el contrario, una región libre de funcionalidad habrá aceptado mutaciones en su secuencia que la habrán ido modificando a lo largo de la evolución de las especies a una tasa que se aproxime a la neutralidad. Una de las grandes ventajas de la secuenciación del genoma humano, junto con el de varias especies de mamíferos, insectos y otros organismos, fue la posibilidad de comparar estos genomas de manera que se pudiesen conocer en mayor detalle los patrones filogenéticos evolutivos a lo largo de los millones de años en los que los distintos genomas se separaron. La comparación de la homología de secuencia entre regiones ortólogas de las especies analizadas frente a la tasa de variación neutra posibilita conocer las regiones de genoma que han estado sometidas a selección negativa en el intervalo temporal de divergencia entre dichas especies.

La primera estimación de conservación a lo largo del genoma entre humano y otra especie surgió tras la publicación de la secuencia de ratón (Chinwalla y cols., 2002), y aproximadamente un 5% de la secuencia del genoma del roedor se encontraba conservada en humano (Chiaromonte y cols., 2003). Poco después, la publicación del genoma del perro permitió extender la comparación del genoma humano con el de otra especie mamífera. Curiosamente, el porcentaje de conservación estimado era de un 5,3% (Lindblad-Toh y cols., 2005), siendo las regiones conservadas en perro y humano pero no en ratón (en los lugares alineables de los 3 genomas) de tan solo un 0,1%. La consistencia entre ambas comparaciones indicaba que la mayoría de *loci* conservados estaban presentes en regiones no codificantes, teniendo en cuenta que el exoma humano tan solo constituye un 1% de la totalidad genoma humano.

Figura 1. 9. Comparación del porcentaje de la secuencia del genoma humano conservada entre varias especies de mamíferos en diferentes estudios. El número de especies analizadas para las estimaciones de conservación se encuentran entre paréntesis. El estudio de Meader y cols (2010) utiliza, además, especies no mamíferas.



Durante los siguientes años, el incremento en el número de especies cuyo genoma era conocido permitió unas mejores estimaciones del coeficiente de selección entre ellos. Desde el año 2003, diversos estudios estimaron la conservación genética de forma más sofisticada (Asthana y cols., 2007; Cooper y cols., 2005; Lindblad-Toh y cols., 2011; Lunter y cols., 2006; Meador y cols., 2010; Parker y cols., 2009; Pollard y cols., 2010; Siepel y cols., 2005; Smith y cols., 2004). Frente al método de Chiaromonte y cols., en el que se recorría el genoma en busca de "ventanas" de selección, surgieron otros basados en la conservación por nucleótido, teniendo

en cuenta las bases flanqueantes. Algunos usaron como comparación de evolución neutral repeticiones ancestrales, presumiblemente no funcionales, originadas a partir de transposones. Otros utilizaron sitios degenerados en codones, muestreo aleatorio del genoma (asumiendo la neutralidad global del mismo) e, incluso, variación en la estructura secundaria del ADN. A pesar de las diferentes metodologías, los valores de selección purificadora entre las especies consideradas mostraban unos valores de conservación entre 3,25% y 12%, lo que era consistente con la presencia de elementos conservados en regiones no codificantes del genoma y, por ello, presumiblemente funcionales (**figura 1.9**).

1.3.3.2. Elementos conservados del genoma humano

Con anterioridad a la publicación de los estudios comparativos entre diferentes genomas, algunos autores ya defendían la búsqueda de regiones reguladoras en humanos mediante el análisis de regiones conservadas en otras especies de mamíferos (Pennacchio y Rubin., 2001). En esta línea, Dermitzakis y cols., mediante la comparación del cromosoma 21 en los genomas de 14 mamíferos, definieron regiones no codificantes conservadas (CNGs) y observaron niveles de selección equiparables o superiores a los esperados en exones, demostrando la previsible arquitectura funcional de genoma no codificante (Dermitzakis y cols., 2002; Dermitzakis y cols., 2003). Un año después, Bejerano y cols. definieron 481 **elementos ultraconservados (UCEs)** como aquellos *loci* de al menos 200 pb íntegramente conservados en los genomas de humano, ratón y rata (Bejerano y cols., 2004), algo que ocurre con una probabilidad menor a 10^{-22} y congregaba un total de 106767 pb. Este primer gran estudio de regiones conservación a lo largo de todo genoma se convertiría en una referencia por sus interesantes conclusiones:

- En primer lugar, se demostró la importancia de la conservación genética en las regiones no codificantes del genoma. De los 481 UCEs, solo 111 solapaban con algún exon conocido, incluyendo regiones UTRs, dejando constancia de que la principal funcionalidad asociada a conservación se encuentra en parte no codificante del genoma.
- Había 93 genes que solapaban con algún UCE en su región exónica, y se denominaron de *tipo I*. El análisis de *pathways* en *Gene Ontology (GO)* reveló que estaban enormemente enriquecidos en las funciones de regulación del splicing y unión a ARN, demostrando el papel de estos genes en procesos de regulación post transcripcional.
- Los 225 genes cercanos al resto de UCEs de regiones no exónicas se denominaron de *tipo II*. Contrariamente a los de tipo I, no se encontraban enriquecidos en ninguna función post transcripcional, sino que se hallaban fuertemente asociados a funciones de unión a ADN y, especialmente, en genes del dominio *Homeobox*. Esto demostraba la

naturaleza de estos genes como reguladores de la transcripción, y particularmente, en procesos del desarrollo básicos. 156 de esos genes flanquean a UCEs no codificantes, situados en regiones intergénicas a más de 100 kb de distancia de ellos, y enriquecidos en genes de desarrollo temprano. Esto demuestra que gran parte de esos elementos reguladores deben tener una función de *enhancer* distal de importancia crítica para regular la expresión de genes claves en el desarrollo precoz de la vida humana.

- Finalmente, de entre los genes solapados o flanqueados por UCEs, había algunos ejemplos particularmente interesantes en el contexto de las enfermedades psiquiátricas y la esquizofrenia. Tal era el caso del gen *ARX*, cercano a los tres UCEs de mayor longitud (> 700 pb) y cuya disrupción está asociada a malformaciones cerebrales, epilepsia y fenotipo autista. Otro caso más interesante, si cabe, es el del *TCF4*, asociado de forma consistente a esquizofrenia de los últimos estudios GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), y que alberga dos UCEs a lo largo de su secuencia.

Conclusiones parecidas se obtuvieron en otros estudios similares donde se analizaron regiones ultraconservadas en genomas de humano, ratón y pez globo (Sandelin y cols., 2004; Woolfe y cols., 2005), siendo la principal característica de los genes cercanos a esas regiones su importancia en procesos del desarrollo y, en particular, que una gran proporción fuesen factores de transcripción. No obstante, la mayor evidencia del papel regulador de estas regiones se obtuvo mediante ensayos *in vivo* de su capacidad de *enhancer*. Se probaron 167 elementos ultraconservados en ensayos de gen *reporter* y el 45% actuaron como tal y, además, específicamente en tejido cerebral y del tubo neural (Pennacchio y cols., 2006). Junto a otros trabajos, el de Pennacchio y cols. constituye la base de datos del VISTA enhancer browser (Visel y cols., 2007; Visel y cols., 2009), donde se almacenan los elementos no codificantes con función de *enhancer* validados experimentalmente. Uno de los dos UCEs del gen *TCF4* es uno de ellos, demostrando la importancia funcional de estos elementos conservados en la expresión de factores de transcripción claves en la patología psiquiátrica como este gen. No obstante, en otro estudio, la delección en ratón de 4 de estos elementos conservados que actúan como *enhancers* en ensayos *in vitro*, no generaba ninguna anomalía perceptible en el desarrollo de los roedores, anunciando que la eliminación de estas secuencias ultraconservadas tampoco tiene porque reflejar funciones críticas en todos los casos (Ahituv y cols., 2007).

Con posterioridad al trabajo de Bejerano, otros usaron modelos más sofisticados para definir regiones de marcada conservación a lo largo de varias especies analizadas. Mediante el programa *PhastCons*, basado en un modelo de Markov oculto en el que se tenía en cuenta la distancia filogenética entre especies y sin la limitación de utilizar intervalos de tamaño determinado, Siepel y cols. detectaron cientos de **elementos de alta conservación (HCEs)** en el

genoma humano (Siepel y cols., 2005). Aplicando el programa a grupos de cinco vertebrados (incluido humano), cuatro insectos, dos *Caenorhabditis* y siete de *Saccharomyces*, independientemente, se observó que la fracción de elementos conservados en la parte no codificante del genoma era mayor cuanto más compleja era la especie. Solo un 42% de los HCEs solapan con exones en vertebrados, mientras que el porcentaje se eleva al 93% en el conjunto de gusanos, levaduras e insectos. A pesar de que fueron miles de elementos detectados (en comparación con solo los 481 de Bejerano y cols.) cubriendo un 4,3% del genoma humano, las propiedades son sorprendentemente similares a las de los UCEs, siendo destacable el hecho de su presencia en regiones alejadas de los genes circundantes por su posible función como reguladores distales de la expresión. Además, es también remarcable la acumulación de estos HCEs en regiones UTR3' asociadas con estructuras secundarias del ARN, dando consistencia a la hipótesis de que un gran número de estos elementos funcionales preservan su secuencia para garantizar un plegamiento específico necesario para una correcta regulación post transcripcional. Por último, es importante señalar que, en consonancia con lo mencionado anteriormente, las categorías funcionales de GO en las que se acumulan estas regiones de forma más significativa fueron las de "unión a ADN", "regulación de la transcripción" (mayormente factores de transcripción) y "desarrollo". Además, otras categorías funcionales de las diecinueve significativas fueron las de "canales iónicos", "remodelación de la cromatina", "neurogénesis", "receptores de glutamato" o "transmisión sináptica", coincidentes con funciones alteradas en esquizofrenia y otros trastornos psiquiátricos reveladas a raíz de estudios genéticos. Esto demuestra que los patrones de selección a lo largo del genoma podrían ser piezas claves en la correcta funcionalidad cerebral, habiendo estado sometidos a una selección purificadora dada su gran importancia. Por ello, las alteraciones de su secuencia conservada a lo largo de millones de años podrían estar asociadas al mal funcionamiento cerebral y desarrollo de sintomatología psiquiátrica.

1.3.3.3. Elementos conservados en mamíferos y acelerados en el linaje humano

No obstante, la búsqueda de elementos de conservación entre varias especies diferentes no asume la posibilidad de cambios en la intensidad de selección a partir de una especie intermedia. Aunque la restricción al cambio en una secuencia conservada a lo largo de una filogenia sea fácilmente identificable, la secuencia que ha ganado o perdido funcionalidad en uno o más linajes, en un momento determinado, será difícilmente distinguible de aquella neutral que siempre ha sido carente de función. De este modo, ni la ganancia de función en un determinado linaje ni una región activa en la especie ancestral que ha perdido función en un linaje pueden ser fácilmente inidentificables (**figura 1.10**). Las estimaciones acerca de estos

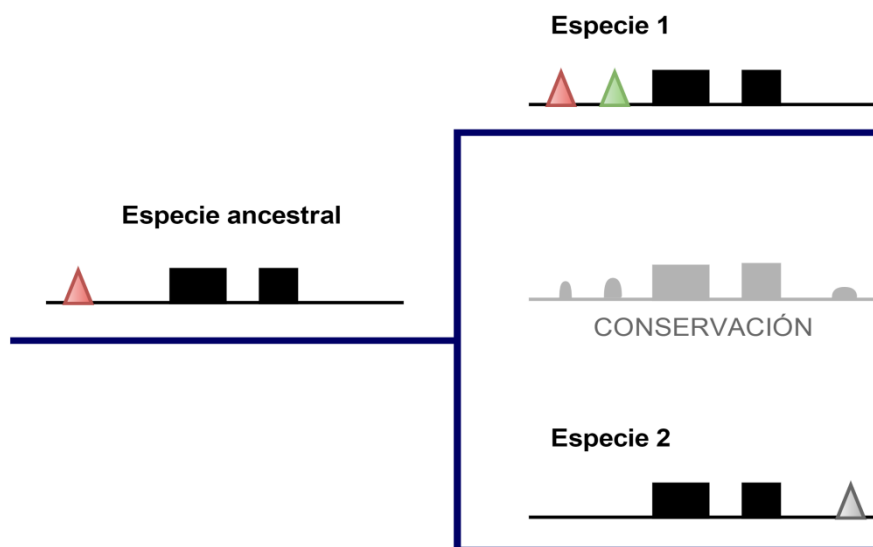
procesos conllevan una subestimación de hasta dos veces el coeficiente de selección detectado (Ponting y Hardison, 2011)

Por tanto, el estudio de regiones conservadas a lo largo del linaje de vertebrados, e incluso mamíferos, permite la búsqueda de patrones de selección estables a lo largo de miles de años. Sin embargo, gracias a la secuenciación del genoma del chimpancé (Sequencing, T. C., & Analysis Consortium, 2005), nuestro mamífero más cercano, la comparación de elementos fuertemente conservados en múltiples especies euterias con cambios recientes del genoma humano respecto a la secuencia de este primate, ofrece la posibilidad de detectar patrones evolutivos propios del linaje humano. Esta divergencia desde nuestro pariente más cercano podría explicar el salto evolutivo en determinadas características propias del intelecto humano, como el desarrollo del lenguaje o el enorme consumo energético del cerebro (Preuss, 2012). Tal es el conocido caso del gen *FOXP2*, el primer gen identificado como responsable del desarrollo del lenguaje, y sometido a selección positiva reciente (Enard y cols., 2002). En el ámbito de la esquizofrenia, una enfermedad propiamente humana con una heredabilidad del 64-80% y asociada a una fertilidad reducida, una hipótesis explicativa de su elevada prevalencia, cercana al 1% en todo el mundo, es la que defiende que varias mutaciones que han originado incrementos en habilidades cognitivas podrían estar favorecidas por selección positiva pero, a la vez, conferir mayor riesgo a esquizofrenia (Crow, 1997). En este sentido, diversos estudios han aportado evidencias empíricas de selección positiva en algunos genes asociados con la enfermedad, aunque es cierto que dichos estudios solo han centrado el foco en un número limitado de genes (Crespi y cols., 2007; Costas y cols., 2009; Lo y cols., 2007).

En regiones reguladoras, la idea de la selección positiva no es algo reciente. Hace ya más de 40 años, se sugirió que la mayor parte de la divergencia entre humanos y chimpancés se encontraría en las regiones no codificantes del genoma (King y Wilson, 1975). Sin embargo, no fue hasta 2006 cuando se realizó la primera búsqueda sistemática a lo largo de todo el genoma de regiones sometidas a selección positiva reciente en humanos o que hubiesen perdido funcionalidad en este linaje (Pollard y cols., 2006a). Las regiones aceleradas en humanos (HARs) son segmentos altamente conservados en especies no humanas pero que experimentan una aceleración significativa en las sustituciones nucleotídicas en el genoma humano. Pollard y cols. demostraron que la presencia de 202 HARs detectados estaba fuertemente asociada a regiones no codificantes del genoma (tan solo un 1,5% de ellos solapaban con exones), pero 88 de ellos presentaban una estructura secundaria determinada mediante modelización *in silico*, lo que sugería la presencia de especies de ARN no codificantes no conocidas de potencial función reguladora. De hecho, el HAR más significativo forma parte de un ARN largo no codificante (lncRNA) que se expresa en el desarrollo de la corteza cerebral humana (Pollard y cols., 2006b). En este sentido, estos autores crearon otro modelo de comparación basándose tanto en la

conservación como en la aceleración respecto a la neutralidad de la variación de secuencia a lo largo de genomas de 36 especies mamíferas diferentes (Pollard y cols., 2010).

Figura 1. 10. Secuencias reguladoras no codificantes no capturadas en estudios de conservación por divergencia entre las especies comparadas. Las regiones codificantes (cuadrados negros) se mantienen en ambas especies, mientras que en regiones reguladoras, algunas pierden su función en la especie 2 (triángulo rojo), y otras ganan función en la especie 1 (triángulo verde) o la especie 2 (triángulo gris).



Otros trabajos, mediante procedimientos distintos e implicando más o menos especies, buscaron elementos conservados que sufrieron aceleración en el linaje humano (Bird y cols., 2007; Bush y cols., 2008; Lindblad-Toh y cols., 2011; Prahbakar y cols., 2006). En el más reciente de ellos, Lindblad-Toh y cols., tras la comparación con 29 especies de mamíferos, describieron la presencia de elementos conservados en mamíferos y acelerados en el linaje humano (mHARs), elementos conservados en mamíferos no primates y acelerados en el linaje primate (PARs) y elementos conservados en primates pero acelerados en humanos (pHARs). Estos últimos demostraban tener un mayor enriquecimiento en funciones cerebrales o del desarrollo como "dirección del axón", "respuesta inmune" o "desarrollo embrionario". Los elementos acelerados de todos los estudios previos fueron recopilados recientemente para definir aproximadamente 1800 elementos acelerados en humanos (HAEs) no solapantes (Kamm y cols., 2013a). El trabajo de Kamm y cols. reveló que *NPAS3* albergaba 14 HAEs diferentes a lo largo de su secuencia, todos ellos no codificantes, siendo con diferencia el gen con el mayor número de estos elementos. Este hecho constituye, además, un soporte para la hipótesis que vincula estas regiones con el desarrollo del cerebro y la esquizofrenia, al ser *NPAS3* un gen asociado a la enfermedad inicialmente a raíz del estudio de una traslocación en su secuencia (Kamnasaran y cols., 2003). En este estudio se demostraba que el papel de *NPAS3* como factor

de transcripción de la familia *bHLH-PAS* (*basic Helix Loop Helix, Period, Aryl hydrocarbon receptor, Single minded*) se veía alterado en una madre y su hija, ambas con esquizofrenia, por afectar la traslocación al dominio *bHLH* de unión a ADN, demostrando la disrupción de su correcta estructura y que la haploinsuficiencia de esta proteína contribuye al desarrollo de la esquizofrenia. Posteriormente, estudios de asociación y familiares confirmaron el papel de *NPAS3* como gen implicado en la enfermedad (Macintyre y cols., 2010; Yu y cols., 2014). Por otra parte, experimentos de *knock-out* de este gen en ratón revelaron fenotipos comparables a los esquizofrénicos (Brunskill y cols., 2005). En el trabajo de Kamm y cols. también se demostró la importancia de estos HAEs como elementos funcionales de regulación de la expresión, al actuar 11 de los 14 HAEs como enhancers transcripcionales durante el desarrollo cerebral en ensayos de *gen reporter* en pez cebra (Kamm y cols., 2013a), un hecho replicado posteriormente en ratón (Kamm y cols., 2013b). La actuación de los elementos acelerados como *enhancers* de la transcripción en otros genes también fue consistentemente hallada en un estudio coetáneo (Capra y cols., 2013), en el que 24 de los 29 HAEs analizados regulaban la expresión en ratón.

Dado que la descripción de elementos acelerados es relativamente reciente y que la mayoría de las variantes genéticas asociadas a esquizofrenia provienen de estudios de poco más de 2 años, pocos estudios han descrito el solapamiento de variantes genéticas en HARs. Sin embargo, un estudio muy reciente (Xu y cols., 2015) demostró un alto enriquecimiento de los haplotipos de señales de GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) en genes separados 100 kb o menos de elementos acelerados pHAR y mHAR descritos previamente (Lindblad-Toh y cols., 2011). Estos genes, presumiblemente regulados por elementos conservados en primates y acelerados en humanos (genes p-HAR), se encontraban fuertemente conservados, lo que establece una diferencia clara entre la conservación de la estructura codificante de genes importantes para el normal desarrollo cerebral y las regiones que regulan su expresión. Respecto a la totalidad de los genes asociados a esquizofrenia, aquellos que presentan elementos HARs cercanos (genes SCZ p-HAR) ocupan lugares más céntricos y conectados en los interactomas de los que forman parte y están estadísticamente enriquecidos en funciones de diferenciación neuronal u otras implicadas en el desarrollo cerebral y de la sinapsis cuando se comparan con la totalidad de genes asociados a esquizofrenia. Este hecho sugiere la interesante posibilidad de que la hipotética selección positiva en estas regiones fuese clave en el salto evolutivo de las funciones cerebrales en el linaje humano respecto al chimpancé, teniendo además una naturaleza puramente reguladora.

El estudio de variación genética en regiones conservadas o aceleradas en el linaje humano también deja, sin embargo, otros estudios previos con resultados negativos (Schanze y cols, 2012; Tolosa y cols., 2008), por lo que las apreciaciones al respecto de un salto evolutivo

que tenga como contrapartida la aparición de trastornos psiquiátricos como la esquizofrenia deben ser tomadas con cautela, y a la espera de estudios de mayor potencia centrados en estas regiones. Es importante señalar que pese a que en el estudio de Xu y cols., los genes cercanos a elementos acelerados en primates (genes PAR) no mostraban ninguna característica de las descritas en los genes pHAR ni los mHAR, otros estudios sugieren que ha habido más cambios en la secuencia y eventos potencialmente selectivos en el linaje del chimpancé que en el nuestro (Bakewell y cols., 2007; Hubisz y cols., 2014; Sequencing, T. C., & Analysis Consortium, 2005), por lo que no seríamos particularmente una especie con mayor aceleración en su secuencia genómica.

Por tanto, en el contexto de la esquizofrenia y de las enfermedades psiquiátricas, es prioritario explorar la naturaleza reguladora del genoma. La búsqueda de elementos conservados o acelerados, que en su mayoría pertenecen a regiones no codificantes, ya ha demostrado la enorme relación existente los genes y las funciones cerebrales que regulan estos elementos y su función demostrada como *enhancers* de la expresión génica.

1.3.4. Regulación de la transcripción

1.3.4.1. Elementos funcionales de la regulación de la expresión génica. Proyecto ENCODE

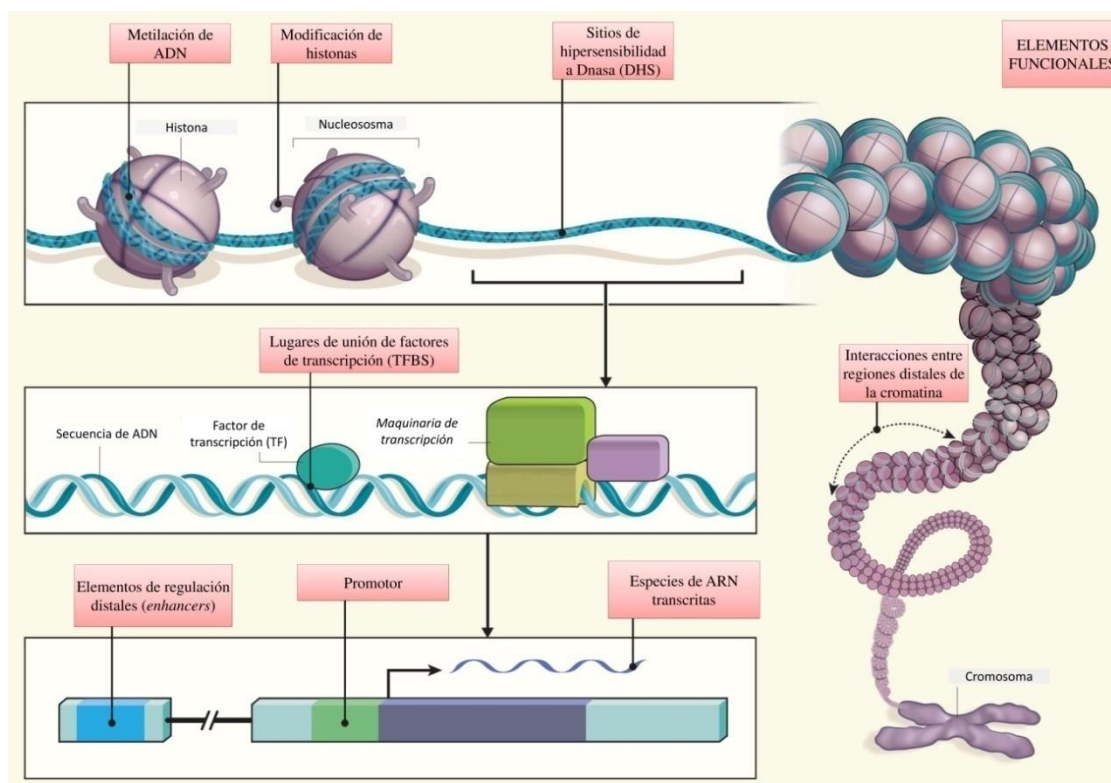
En el año 2003, el NHGRI decide montar el proyecto ENCODE (siglas del inglés *Encyclopedia of DNA Elements*) con el propósito de descubrir y caracterizar todos los **elementos funcionales del genoma**, las especies de ARN transcritas y las modificaciones epigenéticas a lo largo de la cromatina en 147 líneas celulares diferentes. Tras una primera fase piloto hasta 2007, en la que se analiza un 1% del genoma humano a fin de probar y comparar la eficacia, rendimiento y coste de las distintas técnicas utilizables, en 2012 se publican los resultados del proyecto aplicado a la totalidad del genoma en un conjunto de 30 artículos entre los que se reparten los resultados de distintos métodos finalmente empleados. El proyecto ENCODE consiguió caracterizar funcionalmente una gran mayoría del material genético, y al término del mismo se pudieron extraer conclusiones importantes que revelaron una naturaleza reguladora del genoma muy por encima de lo que se creía anteriormente (ENCODE Project Consortium, 2012).

- En contra de lo que se creía previamente, el proyecto ENCODE demostró que la vasta mayoría del genoma (80.4%) presenta actividad bioquímica, bien sea formando parte de alguna especie de ARN transcrita o de algún evento funcional asociado a la cromatina, en al menos una de las líneas celulares estudiadas, demostrando que gran parte del genoma humano podría tener alguna función biológica.

- La clasificación del genoma en estados de la cromatina reveló que casi 400000 regiones a lo largo del genoma y más de 70000 presentaron características funcionales de *enhancers* y promotores en alguna de las líneas celulares estudiadas, algo que se demostró a partir de la extensa batería de experimentos en los que se determinó la unión de factores de transcripción (TF), sitios de hipersensibilidad a DNasa (DHS) o modificaciones histónicas de la cromatina, entre otros.
- Los promotores demostraron ser los principales elementos reguladores de la expresión, la cual se puede predecir a partir de las marcas histónicas y sitios de unión a TF a lo largo de su secuencia. De esta forma, la funcionalidad de estas regiones explica la mayor parte de la variación en la expresión de ARN.
- La mayor parte de los DHS y de los motivos de unión de TF estaban altamente conservados a lo largo de las secuencias de 29 mamíferos (Lindblad-Toh y cols., 2011); y además, mostraron una diversidad en la población humana marcadamente reducida mediante la comparación con la base de datos de los 1000 genomas (1000 Genomes Project Consortium, 2010). No obstante, un elevado porcentaje de estos elementos se encuentra sometido a aceleración en su secuencia, existiendo una gran dispersión respecto a los valores de conservación genética en ellos.
- De forma marcadamente diferente, las secuencias correspondientes a especies de ARN no codificantes, en su mayoría anotadas por primera vez, no mostraron ningún patrón de conservación en mamíferos, pero si una reducida diversidad en población humana. El análisis de 227688 variantes genéticas que segregan en regiones no repetitivas específicas de linaje primate demostró una mayor selección negativa si solapaban con elementos de ENCODE.
- La combinación de de SNP asociados a enfermedades complejas en estudios de GWAS con los elementos funcionales descritos reveló que la mayoría de estas señales de asociación se encuentran dentro o cerca de estas regiones y en su mayoría fuera de las regiones codificantes. En algún caso, además, la enfermedad pudo ser relacionada directamente con un tejido concreto o un factor de transcripción específico.

Los distintos elementos funcionales analizados y las técnicas empleadas se describen en la **figura 1.11** y la **tabla 1.7**, respectivamente. La combinación de estas técnicas con el estudio del transcriptoma en diversas líneas celulares ha permitido conocer en mayor profundidad la naturaleza de las regiones directamente implicadas en la regulación de la expresión génica, como los *enhancers* y los promotores, y las estructuras de la cromatina y modificaciones epigenéticas características de ellas.

Figura 1. 11. Elementos funcionales del genoma descritos en el proyecto ENCODE a lo largo de diferentes líneas celulares (Figura modificada de Ecker y cols., 2012)



En efecto, uno de los resultados destacables del proyecto ENCODE es la demostración de que los promotores, que se extienden simétricamente a ambos lados del TSS abarcando un promedio de 2,5 kb (Birney y cols., 2007), explican la mayor parte de la variabilidad en los niveles de expresión génica y que éstos se correlaciona con marcas de acetilación y metilación en las histonas (r de Pearson = 0.9) y, en menor medida, con la presencia de factores de transcripción (r de Pearson = 0.81) (Dong y cols., 2012; Sanyal y cols., 2012; Thurman y cols., 2012). En general, las modificaciones activadoras (H3K27ac, H3K9ac, H3K4me3 o H3K4me) explican la mayor parte de esta variación, aunque para un grupo reducido de elementos promotores, las marcas represivas (H3K27me3 o H3K9me3) son particularmente importantes para predecir los niveles de expresión. El análisis de las 12 modificaciones de histonas de marcada importancia a lo largo del genoma de 46 líneas celulares diferentes demostró la importancia de las mismas a la hora de caracterizar la función de los diferentes elementos funcionales (**tabla 1.8**), además de una marcada variabilidad entre distintos tejidos y un elevado dinamismo a lo largo de la vida (ENCODE Project Consortium, 2012).

Tabla 1. 7. Descripción de los elementos reguladores estudiados más representativos, las principales técnicas empleadas y los resultados al respecto del proyecto ENCODE (Djebali y cols., 2012; Gerstein y cols., 2012; Neph y cols., 2012; Sanyal y cols., 2012; Thurman y cols., 2012).

Elementos/regiones	Técnica principal	Fundamento	Nº de líneas celulares	Resultados principales
Unión de factores de transcripción	ChIP-seq	Imunoprecipitación de la cromatina mediante el anticuerpo adecuado específico del TF de interés (ChIP) seguido de ultrasecuenciación de la región unida al él.	72	<ul style="list-style-type: none"> - 119 TF se encuentran unidos a 231 Mb del genoma (8,1%). - La unión de TF se produce de forma coordinada junto con otras proteínas.
Modificaciones histónicas	ChIP-seq	Imunoprecipitación de la cromatina mediante el anticuerpo adecuado específico de la modificación química de interés (ChIP) seguido de ultrasecuenciación de la región unida al ella.	46	<ul style="list-style-type: none"> - Los patrones de 12 modificaciones de histonas varían mucho entre líneas celulares diferentes y acorde a variaciones en los niveles de expresión. - La integración, únicamente, de las diferentes modificaciones permite caracterizar las funciones de las regiones afectadas.
Accesibilidad de la cromatina	Dnasa-seq	Secuenciación de las regiones accesibles de la cromatina (DHS) cortadas por la enzima DNasa I.	125	<ul style="list-style-type: none"> - 2.9 millones de DHS no solapantes fueron mapeados a lo largo del genoma. - 970000 fueron específicos de línea celular. - Solo 3629 se encuentran en todas las líneas.
	FAIRE-seq	Secuenciación de las regiones accesibles de la cromatina (DHS no ligadas a proteínas histónicas tras tratamiento previo con formaldehído y fragmentación del genoma mediante sonicación.	25	<ul style="list-style-type: none"> - 5% de los DHS se encuentran en promotores, siendo la mayoría distales. - Los DHS de elementos distales son más específicos de línea celular. - 98.5% de los TFBS ocupaban regiones de DHS.
Huellas o <i>footprints</i> de DNasa	footprinting digital	Análisis de las regiones DHS a nivel de nucleótido para localizar la secuencia del motivo de unión del TF.	41	<ul style="list-style-type: none"> - Se encontraron 8.4 millones de <i>footprints</i> (Neph y cols., 2012). - Los motivos de unión se almacenaron en la base de datos <i>Factorbook</i> (Wang y cols., 2013).
Metilación de ADN	RRBS (<i>Reduced representation bisulphite sequencing</i>)	Secuenciación de regiones ricas en CpG tras tratamiento previo con bisulfito, convirtiendo citosinas no metiladas en uracilos	82	<ul style="list-style-type: none"> - Se describieron patrones de metilación a lo largo de 1.2 millones de dinucleótidos CpG. - Generalmente, la metilación de la cromatina conlleva una limitación en su accesibilidad. - elevada variabilidad en tejidos diferentes.
Interracciones de la cromatina	Captura de la Conformación de los Cromosomas (3C - 5C)	Captura de interacciones de elementos distales con TSS mediante digestión con enzimas de restricción y secuenciación de las regiones que interaccionan	4	<ul style="list-style-type: none"> - Alta correlación entre la interacción de TSS de los genes y <i>enhancers</i> y la expresión de genes que regulan. - De media, 3.9 elementos distales interactúan con un mismo TSS. - De media, 2.3 TSS interactúan con un mismo elemento distal.

Tabla 1. 8. Modificaciones químicas de histonas estudiadas en el proyecto Encode (ENCODE Project Consortium, 2012; Kouzarides, 2007; Zhou y cols., 2011). La simbología de las modificaciones representa, de forma secuencial, la familia de las proteínas histónicas alterada (H2A, H2B, H3 o H4), la posición del residuo de lisina modificado (K4, K9, K20, K27, K36 o K79) y el tipo de modificación química (me1 = metilación; me2 = dimetilación; me3 = trimetilación; ac = acetilación). En el caso de H2A.Z, ésta representa una variante de la proteína H2A.

Modificación histónica	Señal característica	Función
H2A.Z	Pico	Elementos reguladores con cromatina dinámica
H3K4me1	Pico/región	Principal marca de <i>enhancers</i> y otros elementos de regulación distal, pero también <i>downstream</i> de algunos TSS
H3K4me2	Pico	Promotores y <i>enhancers</i>
H3K4me3	Pico	Principal marca reguladora de promotores en TSS
H3K9ac	Pico	Marca activa de elementos de regulación, preferencialmente promotores
H3K9me1	Región	Localización en extremos 5' de genes
H3K9me3	Pico/región	Marca represiva de heterocromatina y elementos repetitivos
H3K27ac	Pico	Marca activa de elementos de regulación, distingue entre <i>enhancers</i> y promotores activos de los inactivos
H3K27me3	Región	Marca represiva característica del complejo <i>Polycomb</i> y asociada al silenciamiento de genes del desarrollo
H3K36me3	Región	Marca de elementos transcritos preferentemente en regiones en 3'
H3K79me2	Región	Marca de elementos transcritos preferentemente en regiones en 5'
H4K20me1	Región	Localización en extremos 5' de genes

Dados los elevados niveles de conservación genética, y la importancia funcional que se consolida en el proyecto ENCODE, el estudio de la variación genética en los elementos funcionales descritos puede ser un paso importante a la hora de dilucidar el papel de la genética en los procesos de regulación de la expresión génica. En este sentido, un último objetivo del proyecto ENCODE fue la interpretación las señales de asociación de GWAS y de las variantes raras de estudios de secuenciación en regiones no exónicas, lo que reveló relaciones importantes entre la variación y los elementos funcionales previamente descritos.

- Dado que el 88% de las variantes asociadas a enfermedades complejas son intrónicas o intergénicas (Hindorf y cols., 2009), se examinaron 4860 asociaciones para 4492 SNP de la base de datos del NHGRI, encontrándose que un 12% de ellos solapaban con sitios de unión de factores de transcripción y un 34%, regiones de DHS.
- Al extender las señales de asociación a los intervalos en LD, los valores de solapamiento ascienden a un 31% y un 71% en el caso de unión a factor de transcripción o DHS, respectivamente. Bajo esta perspectiva, una clara mayoría de las

señales de asociación pueden acarrear, en mayor o menor medida, una disrupción en las vías de regulación afectadas.

- Los SNP analizados también se encuentran particularmente asociados a regiones con funcionalidad de *enhancer* o sitios cercanos a TSS, demostrando el importante papel de estos elementos en las vías de regulación genética afectadas en las principales enfermedades comunes.

1.3.4.2. Estados epigenómicos específicos de tejido. Proyecto Roadmap.

Una de las conclusiones que se extrajo del análisis combinado de los elementos funcionales de ENCODE y de variantes genéticas asociadas a enfermedades fue que existía una marcada correspondencia entre el tipo celular y el fenotipo analizado, algo en consonancia con las numerosas evidencias surgidas a lo largo de la descripción de los elementos funcionales que apuntaban a una especificidad hasta de línea celular o tejido, como en los *enhancers*, patrones de metilación y modificaciones histónicas.

En este sentido, una de las limitaciones del proyecto ENCODE fue que, pese a la extensa caracterización del transcriptoma y los elementos reguladores que determinan su expresión, la aplicación a regiones concretas del individuo adulto como las distintas áreas cerebrales o pulmonares estaba limitada por la escasez de líneas celulares de algunos tejidos de elevado interés. En el caso de la esquizofrenia y el ámbito del cerebro humano, solo dos líneas celulares (que en muchas ocasiones no estaban caracterizadas para todos los experimentos) pertenecían a cerebro adulto: SK-N-SH y NH-A.

Por ello, una de los propósitos de los propósitos de NIH fue impulsar un proyecto, el proyecto Roadmap (Kundaje y cols., 2015), que ampliase el espectro de análisis en 111 nuevos epigenomas de prácticamente todos los tejidos y con representación de varias áreas o líneas celulares en cada uno. Esa nueva batería de epigenomas se sumó a los de las 16 líneas más caracterizadas existentes en la base ENCODE para dar lugar a 127 epigenomas con información de DHS, metilación, transcriptoma y modificaciones histónicas. Algunos resultados destacables de la publicación inicial del proyecto fueron los siguientes:

- La combinación de las modificaciones histónicas presenta una alta variación en los patrones de metilación y accesibilidad, prediciendo, además, cambios en los niveles de ARN independientemente de la metilación o los niveles accesibilidad.
- Se detectó la presencia de dominios cromosómicos de gran tamaño, en los cuales hay marcas características y actividad y densidad génica marcadamente diferentes.

- Una media del 5% del epigenoma de los distintos tejidos presenta marcas de promotores y *enhancers* con un nivel de conservación dos veces el nivel relativo a la neutralidad, lo cual es consistente con los patrones de conservación del proyecto ENCODE y con los elementos conservados descritos con anterioridad. Los elementos *enhancers* muestran una mayor especificidad que los promotores, y ambos se hallan asociados a marcas de regulación y dominios de *footprinting*.
- Por último, las variantes genéticas asociadas en enfermedades complejas en GWAS se pueden asociar en multitud de casos a marcas epigenéticas que tienen lugar en la línea celular o tejido involucrado en la patología, lo que viene a consolidar los primeros indicios al respecto en las células estudiadas en ENCODE. En el contexto de la esquizofrenia, gracias a los datos en tejido cerebral aportados por el proyecto Roadmap, las señales de asociación del último GWAS se encuentran enriquecidas en regiones activas (H3K27ac) en los tejidos cerebrales del giro angular y el lóbulo frontal medial (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

Además, entre los principales volúmenes de datos aportados por el proyecto Roadmap se encuentran los estados de la cromatina específicos de los tejidos estudiados a partir de datos epigenéticos. Basándose en un modelo Markov oculto (HMM), se predijeron 15 estados de la cromatina a partir de 5 marcas histónicas en los 127 epigenomas de los que se disponían, revelándose que 68% de cada epigenoma, de media, era quiescente y *a priori*, inactivo. Sin embargo, la elevada divergencia entre estados de distintos epigenomas y su alta especificidad de tejido puede explicar este alto número de regiones quiescentes en cada caso. De los 15 estados descritos, 8 son activos y 7 represivos (**tabla1.9**).

Recientemente, una sección del consorcio del ENCODE, denominada Psych ENCODE, se ha puesto como objetivo caracterizar en mayor profundidad la complejidad de las funciones reguladoras en el cerebro (Akbarian y cols., 2015). Para ello, haciendo uso de las bases de datos propias, de ENCODE, Roadmap, GTEx y otras relacionadas, este consorcio pretende caracterizar los estados funcionales del genoma desde un contexto específico sin precedentes, con el fin de desentrañar las claves de la regulación genética en el neurodesarrollo, la evolución del cerebro humano y la predisposición a padecer trastornos psiquiátricos mediada por la presencia de variantes genéticas concretas. Para este último objetivo, también se dispondrá de datos del epigenomas de individuos enfermos y controles sanos. De este modo, se pretende hacer un estudio comparativo y detallado de los perfiles espacio-temporales de expresión a lo largo del desarrollo del cerebro humano, y así conocer en mayor detalle las variantes asociadas a la enfermedad psiquiátrica (Akbarian y cols., 2015).

Tabla 1. 9. Descripción de los 15 estados funcionales de la cromatina caracterizados en Kundaje y cols. (2015)

Estado	Descripción	Marcas histónicas	% cobertura
TssA	TSS activo	H3K4me3	0.70%
TssAFlnk	Flancos de TSS activo	H3K4me3, H3K4me1	0.50%
TxFlnk	Tránsito en 5' y 3'	H3K4me3, H3K4me1, H3K36me3	0.10%
Tx	Elevada transcripción	H3K36me3	3.60%
TxWk	Débil transcripción	H3K36me3 (débil)	11.60%
EnhG	<i>Enhancers</i> en gen	H3K4me1, H3K36me3	0.40%
Enh	<i>Enhancers</i>	H3K4me1	2.80%
ZNF/Rpts	Genes de ZNF y repeticiones	H3K4me3, H3K36me3, H3K9me3	0.20%
Het	Heterocromatina	H3K9me3	2.60%
TssBiv	TSS Bivalente	H3K4me3, H3K27me3	0.10%
BivFlnk	Flanco de TSS/ <i>enhancer</i> bivalente	H3K4me3, H3K4me1, H3K27me3	0.10%
EnhBiv	<i>Enhancer</i> bivalente	H3K4me1, H3K27me3	0.10%
ReprPC	Represor Polycomb	H3K27me3	1.20%
ReprPCWk	Represor Polycomb débil	H3K27me3 (débil)	8.30%
Quies	Quiescente	-	67.80%

1.3.4.3. Estudio de la variación genética en elementos reguladores

Dado el gran avance que supusieron los proyectos anteriormente descritos en la caracterización funcional del genoma no codificante, y quedando demostrada la influencia de las regiones reguladoras en el desarrollo de las enfermedades complejas, en los últimos años se han publicado diversas herramientas de acceso libre *online* para poder anotar las variantes genéticas en base a la información funcional generada por el proyecto ENCODE y relacionados. De este modo, y de forma análoga a las herramientas existentes para la predicción *in silico* de la funcionalidad de una variante codificante como PolyPhen (Adzhubei y cols., 2010) o SIFT (Ng y Henikoff, 2003), actualmente se puede anotar el impacto funcional en la regulación de la expresión génica de variantes no codificantes gracias a herramientas como RegulomeDB (**tabla 1.10**) (Boyle y cols., 2012) o HaploReg (Ward y Kellis, 2012).

En el ámbito de la esquizofrenia y las enfermedades psiquiátricas, estudios recientes, valiéndose de la información acerca de elementos reguladores descritos, han empezado a encontrar evidencias que vinculan la variación genética en estas regiones con el desarrollo del trastorno psiquiátrico (Duan y cols., 2014; Roussos y cols., 2014).

El estudio de Roussos y cols. revela un enriquecimiento significativo de las variantes asociadas a esquizofrenia del último GWAS por aquel entonces (Ripke y cols., 2013a) en

elementos reguladores de la expresión en *cis* (CREs). Además, se encuentra un solapamiento significativo entre SNP que afectan a la expresión (eSNP) y esos elementos CREs (creSNP), en consonancia con estudios previos que revelaban una asociación entre alelos de susceptibilidad y aquellos que afectan a la expresión en cerebro humano (Richards y cols., 2012). Roussos y cols. encontraron la mayor asociación con esquizofrenia en los creSNP de *enhancers* en cerebro fetal y adulto, confirmando hallazgos previos que ya señalaban la importancia crítica de los *enhancers* en la organización de la transcripción durante el neurodesarrollo (Wenger y cols., 2013). En la misma línea que el trabajo de McVicker y cols. (2013), Roussos y cols., también demostraron la interacción en *trans* entre elementos reguladores distales y proximales en el locus del gen *CACNA1C*, sugiriendo la existencia de redes de mecanismos bioquímicos de regulación de elevada complejidad.

Tabla 1. 10. Categorías funcionales de RegulomeDB en función de las evidencias de alteración de la transcripción a partir de datos de ENCODE.

Descripción	Categoría	requisitos funcionales
<i>Muy probable que afecte a la unión de factores de transcripción y relacionado con variación en la expresión</i>	1a	eQTL + unión a TF + motivo del mismo TF + DHS + <i>Footprint</i> del mismo TF
	1b	eQTL + unión a TF + Cualquier motivo de unión + DHS + <i>Footprint</i>
	1c	eQTL + unión a TF + motivo del mismo TF + DHS
	1d	eQTL + unión a TF + Cualquier motivo de unión + DHS
	1e	eQTL + unión a TF + motivo del mismo TF
	1f	eQTL + Unión a TF o DHS
<i>Muy probable que afecte a la unión de factores de transcripción</i>	2a	Unión a TF + motivo del mismo TF + DHS + <i>Footprint</i>
	2b	Unión a TF + Cualquier motivo de unión + DHS + <i>Footprint</i>
	2c	Unión a TF + motivo del mismo TF + DHS
<i>Menos probable que afecte a la unión de factores de transcripción</i>	3a	Unión a TF + Cualquier motivo de unión + DHS
	3b	Unión a TF + motivo del mismo TF
<i>Evidencia de unión mínima</i>	4	Unión a TF + DHS
	5	Unión a TF o DHS
	6	Cualquier motivo de unión

Por otra parte, el estudio de Duan y cols., centrado en la búsqueda de variantes raras de predisposición que afectasen a la expresión del gen *MIR137*, uno de los genes más fuertemente asociados a esquizofrenia en estudios GWAS pero sin resultados similares en los estudios de exoma, se basó en la secuenciación de casi 7 kb de regiones reguladoras alrededor de dicho gen en 2610 casos y 2611 controles. Para la selección de regiones reguladoras, los autores se basan en las regiones DHS de las células neuronales de ENCODE (SK-N-SH y NH-A) y en cerebro fetal, anotando estas secuencias seleccionadas en función de las marcas de histonas (tabla 8)

como promotores (presencia de H3K4me3), *enhancers* (presencia de la marca H3K4me1) e *insulators* (cuando presentan la marca represiva de CTCF, que anula el efecto del *enhancer*). El análisis filogenético de las regiones reguladoras seleccionadas revelaba unos niveles de variación genética similares a los esperados para variantes *missense* o *frameshift*, por lo que claramente eran regiones sometidas a selección purificadora.

Los resultados del estudio de Duan y cols. mostraron una acumulación significativa de variantes en regiones activas de promotores y *enhancers*, mientras que la significación desaparecía si se incluían las variantes en *insulators* (enriquecidas, además, en controles aunque no significativamente). Estos datos revelaron la importancia de analizar la acumulación de variantes raras en regiones reguladoras en función de los elementos funcionales afectados, ya que muchas veces pueden tener tendencias contrapuestas, como en el caso. Sin la información de la caracterización funcional necesaria para dividir elementos activos y represores de la transcripción, no se habría podido desvelar el riesgo asociado a las variantes en regiones reguladoras activas. El estudio permitió, además, tras la identificación de una variante en 11 casos y 0 controles y su posterior genotipado en miles de casos, demostrar la asociación de la misma con la esquizofrenia, el impacto que provoca en la unión a TF y su funcionalidad bioquímica como enhancer específica de tejido. Por tanto, el trabajo de Duan y cols. representa un ejemplo de cómo algunos genes asociados a esquizofrenia, pero sin variaciones en su secuencia exónica, pueden ejercer su predisposición a través de variantes reguladoras, menos agresivas que las codificantes, y que concordarían en mayor medida con un fenotipo común en la población general.

Por otra parte, las modificaciones epigenéticas que describen los distintos estados funcionales de la cromatina han demostrado ser un factor de predisposición para el desarrollo de la enfermedad mental. Como ejemplo, el análisis de los genes con diferencias significativas de metilación en humano respecto a chimpancé demuestra una clara asociación de los mismos con categorías funcionales propias de enfermedades características del linaje humano como "trastornos neurológicos", "trastornos psicológicos" y "cáncer" (Zeng y cols., 2012). Sin embargo, a pesar de las numerosas evidencias de que los estados epigenéticos alterados estén relacionados con la vulnerabilidad a enfermedades humanas, pocos estudios han centrado su investigación en la influencia que ejerce la propia secuencia genética en que existan estos patrones alterados que condicionan la expresión y activación o silenciamiento de diversos genes. En esta línea, algunos trabajos recientes sugieren que una gran parte de esas modificaciones histónicas u otras marcas epigenéticas que regulan los niveles de expresión tienen un origen principalmente genético (Kilpinen y cols., 2013; McVicker y cols., 2013), y que, por tanto, demuestran la importancia de estudiar, junto con el epigenoma y los niveles de expresión, la variación genética que puede subyacer a estas alteraciones. Prueba de ellos es que

las variantes genéticas que afectan a elementos de regulación distales se relacionan tanto con cambios en la expresión como con modificaciones de la cromatina en las regiones promotoras de los genes que regulan (McVicker y cols., 2013).

En este sentido, los análisis más recientes de los resultados de GWAS demuestran que los *loci* asociados a enfermedades psiquiátricas están enriquecidos en genes implicados en la función de metilación de histonas, especialmente de H3K4 (Network, T. & Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015). En la misma línea, otro estudio reciente en tríos de esquizofrenia demuestra la presencia significativa de mutaciones en genes asociados a enfermedades mentales y cuyas funciones incluyen el remodelado de la cromatina mediante modificaciones histónicas (McCarthy y cols, 2014). Estos resultados demuestran la importancia de considerar el origen genético de la alteración de los estados epigenómicos y, por ende, de los niveles de expresión de los genes que regulan.

Por tanto, y a pesar de lo reciente de la descripción funcional del genoma no codificante, hay numerosas evidencias que justifican el estudio de la variación genética en las regiones reguladoras. Los escasos estudios al respecto, ya han demostrado la influencia que ejercen las variantes genéticas reguladoras en la predisposición a esquizofrenia, mediada por alteraciones en la expresión de tejidos previamente relacionados como la corteza prefrontal. Estas conclusiones se hacen más visibles todavía en las interpretaciones de los resultados de los GWAS más importantes, donde, sin embargo, la explicación de varianza genética es insuficiente aludiendo a esta variación común. La hipótesis enfermedad común - variación rara en el contexto de las regiones reguladoras, muy poco estudiada hasta el momento, podría explicar parte de esa heredabilidad perdida característica de la esquizofrenia. Por ello, son necesarios estudios de la variación genética rara en regiones reguladoras (especialmente promotores y *enhancers*), y seleccionadas en base a marcas epigenéticas concretas, criterios de conservación o posible selección, para tratar cubrir todo el espectro de variación y conocer en profundidad los factores genéticos de predisposición al desarrollo de la esquizofrenia.



2. JUSTIFICACIÓN, HIPÓTESIS Y OBJETIVOS



JUSTIFICACIÓN

La esquizofrenia es una enfermedad compleja con una alta heredabilidad estimada, lo cual justifica su estudio mediante la búsqueda de variantes genéticas de susceptibilidad. Los estudios de GWAS han identificado numerosos *loci* implicados en la etiología de la esquizofrenia, así como rutas biológicas de mayor susceptibilidad. No obstante, a pesar del empleo de varios miles de muestras, estos estudios no han conseguido explicar la mayor parte de la heredabilidad esperada para la enfermedad.

Los estudios de secuenciación masiva se han restringido al análisis de la variación rara en regiones codificantes. A pesar de haber destacado rutas biológicas implicadas en la etiología de la esquizofrenia, como los genes de la densidad postsináptica, sus resultados revelan una implicación de éstas menor de la esperada.

El estudio de la conservación genética ha permitido desvelar la naturaleza de elementos conservados en el genoma humano, en su mayoría de naturaleza no codificantes, y cuya implicación en procesos de regulación de la expresión génica durante el desarrollo y, especialmente, neurodesarrollo, ya ha sido demostrada. Igualmente, elementos conservados en mamíferos, pero que han experimentado una aceleración en los cambios nucleotídicos de su secuencia en el linaje humano, han demostrado estar implicados en funciones de regulación de la expresión génica de genes implicados en procesos del neurodesarrollo.

Diversos estudios y consorcios, como el proyecto Encode o el proyecto Roadmap han permitido conocer en mayor profundidad la naturaleza de las regiones reguladoras a lo largo del genoma, la interpretación funcional de las variantes genéticas en ellas y las diferencias de expresión génica en las diferentes líneas celulares. Otros, incluso, como el proyecto BrainSpan, se han centrado en el estudio de la expresión génica en cerebro, de forma específica.

Las regiones promotoras de los genes siempre han sido consideradas los principales elementos de regulación de la expresión génica. Al margen de lo ventajoso de su análisis por ser regiones fácilmente localizables, el estudio estructural de los promotores han revelado con mayor exactitud la arquitectura de los mismos, e igualmente demostrado su papel como los elementos que explican la mayor parte de la variabilidad en la expresión génica, en un contexto que implica tanto modificaciones genéticas como epigenéticas.

Por último, la tecnología de ultrasecuenciación aplicada al estudio del transcriptoma (RNA-seq) ha permitido la posibilidad del análisis combinado de las variaciones en la expresión génica *in situ* y de la variación genética presente en las especies expresadas en el tejido de interés.

HIPÓTESIS

Es probable que la mayor parte de la variación genética de susceptibilidad a esquizofrenia sea de frecuencia baja en la población, ajustándose esta enfermedad a un modelo de variante rara - enfermedad común.

La mayor parte de esas variantes genéticas raras de susceptibilidad podrían encontrarse en las regiones no codificantes, influyendo de forma notable en las alteraciones de la expresión génica de aquellas especies implicadas en procesos biológicos de relevancia para el desarrollo de la esquizofrenia.

La variación genética presente en elementos conservados podría aumentar el riesgo a la aparición de trastornos del neurodesarrollo como la esquizofrenia, en base a una potente selección purificadora a lo largo de la evolución. Igualmente, las regiones conservadas que hubiesen experimentado una aceleración en su secuencia en alguna etapa del desarrollo del linaje humano podrían albergar variación genética de riesgo a esquizofrenia, dada la demostrada funcionalidad reguladora de las mismas.

El estudio de la variación genética en las regiones promotoras de genes de susceptibilidad a esquizofrenia, basándose en los datos epigenéticos y de expresión generados por recientes trabajos y consorcios asociados, es, probablemente, un paso imprescindible para el estudio de la etiología genética de la enfermedad en el contexto de la regulación de la expresión génica.

La existencia de varias CNV de riesgo a esquizofrenia sugiere la importancia de los niveles de expresión de múltiples genes afectados como un factor de susceptibilidad al desarrollo de esquizofrenia. La variación en los niveles de expresión de genes que codifican para factores de transcripción (que, a su vez, regulan un alto número de genes), podría tener consecuencias todavía mayores.

Finalmente, el análisis de las variantes genéticas presentes en el ARN de cerebro humano, en relación con los cambios en los niveles de expresión, podría representar una estrategia muy ventajosa para conocer las variantes de susceptibilidad a esquizofrenia en una relación directa con los cambios en la expresión génica.

OBJETIVOS

El objetivo central de los trabajos que conforman el proyecto de tesis aquí presentado fue la identificación de factores genéticos de susceptibilidad a esquizofrenia bajo la hipótesis de una implicación de los mismos en la alteración de la expresión génica, así como el estudio en profundidad de los cambios en los niveles de expresión a nivel cerebral en relación con las variantes identificadas.

En función de las estrategias y del modo en que se abordaron estos objetivos en los cuatro trabajos que componen este proyecto, se puede establecer una división del mismo en tres bloques:

- En primer lugar, se estudió la variación genética en regiones del genoma sometidas a patrones de evolución extremos. Este bloque consistió de dos estudios:
 - Por un lado, se buscaron variantes raras de susceptibilidad en uno de los genes candidatos a esquizofrenia con mayor número de elementos ultraconservados en su secuencia: *TCF4*
 - Por otro lado, se analizó la variación genética presente en el gen con mayor número de elementos acelerados en humanos y, además, relacionado con la etiología de la esquizofrenia y alteraciones durante el neurodesarrollo: *NPAS3*.
- En segundo lugar, en lo que comprendió el trabajo de mayor dedicación en el presente proyecto de tesis, se estudió de forma sistemática la variación genética rara de susceptibilidad en las regiones promotoras de genes candidatos a esquizofrenia recogidos por evidencias en estudios de CNV y de GWAS, y mediante el empleo de datos de regulación generados por recientes proyectos como Roadmap, Encode o BrainSpan.
- En tercer y último lugar, se estudiaron las alteraciones de expresión en cerebro humano en esquizofrenia a partir del análisis comparado de transcriptoma de individuos afectos y controles sanos, así como las variantes de susceptibilidad a esquizofrenia relacionadas con las alteraciones de expresión de aquellos genes altamente expresados en las muestras estudiadas.





3. MATERIALES Y MÉTODOS



3.1. MUESTRAS ANALIZADAS

3.1.1. Información de las muestras empleadas

3.1.1.1. Muestras de ADN.

Las muestras empleadas en los trabajos I-III fueron obtenidas de la colección total de 580 pacientes de esquizofrenia y 577 controles. Los estudios realizados fueron aprobados por el Comité Ético de Investigación Clínica de Galicia, organismo independiente acreditado por la Consellería de Sanidade de la Xunta de Galicia, y se llevaron a cabo en conformidad con la Declaración de Helsinki. Una descripción adicional de los sujetos se presenta más detalladamente en el estudio de Carrera y cols. (Carrera y cols., 2012).

En todos los casos, los individuos fueron diagnosticados por psiquiatras experimentados que obraron en base a los criterios establecidos en el manual de psiquiatría DSM-IV. Todas las muestras empleadas procedieron de donaciones de sangre periférica por parte de enfermos, previo consentimiento informado, de los centros hospitalarios del área de Santiago de Compostela, y todas ellas son de origen gallego. Como criterios de inclusión, se estableció que todos los individuos afectos debían haber tenido el diagnóstico de esquizofrenia al menos en los últimos dos años y que, además, no tuvieran relación de parentesco alguna con otros ya incluidos previamente en la base de datos. En este sentido, la mayor parte de las muestras se genotiparon previamente, en muchos SNP, a lo largo del genoma, lo que permitió la detección de las relaciones de parentesco entre ellas y la estratificación, con la consecuente eliminación de las muestras afectadas (Carrera y cols., 2012). Brevemente, tras los procedimientos de control de calidad, se utilizaron un total de 5100 SNP *missense* comunes para detectar la relación críptica por cálculos de identidad por estado (del inglés *identity by state*) utilizando el software PLINK (Purcell y cols., 2007), y un total de 130 marcadores informativos de ancestralidad fueron empleados para la detección de la estructura poblacional mediante el *software* Structure v2.3.1 (Pritchard y cols., 2000). Los casos de trastorno esquizofreniforme y esquizoafectivo fueron, igualmente, excluidos del total de muestras aquí presentes.

Los controles empleados, de carácter anónimo, se recogieron del Centro de Transfusión de Sangre de Galicia en espacios de tiempo lo suficientemente cortos como para asegurarse que no se incluyeran por repetido. Además, no se entregó remuneración económica en ningún caso, a fin de no incurrir en algún sesgo de selección de grupos sociales más desfavorecidos.

En el **Estudio I**, relativo a la búsqueda de variantes genéticas de susceptibilidad en elementos reguladores ultraconservados de *TCF4*, se emplearon 528 casos esquizofrénicos y 270 controles de la base de datos de muestras, anteriormente descrita. La media de edad de las

muestras analizadas fue comparable entre ambos grupos, siendo 45,5 en casos y 41,5 en controles. Respecto al sexo, en los controles hubo un 50% de varones por un 67% en casos esquizofrénicos.

En el **Estudio II**, relativo a la búsqueda de variantes genéticas de susceptibilidad en elementos acelerados en humanos a lo largo de las regiones reguladoras de *NPAS3*, se emplearon 538 casos esquizofrénicos y 539 controles de la base de datos. La media de edad de las muestras analizadas en este caso fue 49,1 en casos y 40,9 en controles, mientras que, en lo relativo al sexo, un 57% de los primeros fueron varones por un 49,5% en los segundos.

En el **Estudio III**, centrado en la búsqueda de variantes genéticas de susceptibilidad en las regiones promotoras de genes relacionados con la esquizofrenia a partir de estudios de GWAS y de CNV, se utilizaron un total de 516 pacientes esquizofrénicos de la base de datos descrita, así como 516 controles sanos, de origen gallego. La media de edad y el porcentaje de varones en los casos fue de 49,5 y 67%, respectivamente, mientras que en controles fue de 57 años de edad media y de 50,6% de varones.

3.1.1.2. Muestras de ARN.

En el último trabajo del presente proyecto de tesis, el **Estudio IV**, se utilizó una colección de muestras *post mortem* de ARN de la región cerebral del giro cingulado procedentes de 31 casos esquizofrénicos y 26 controles sanos. Las muestras fueron obtenidas de la *colección Array* del repositorio de tejidos del *Stanley Medical Research Institute* (SMRI). Estas muestras analizadas incluyeron información sobre la exposición estimada de por vida a los antipsicóticos, pH cerebral y otros datos puestos a disposición del investigador por el SMRI. El pH cerebral se determinó de tejido homogeneizado de la corteza occipital, y la integridad y pureza del ARN se determinaron con un *Agilent 2100 Bioanalyzer* (Agilent technologies, Santa Clara, CA).

Las muestras de ARN se recogieron, previo consentimiento informado de los familiares más cercanos, por examinadores médicos colaboradores del SMRI entre enero de 1995 y junio de 2002. Los especímenes fueron recolectados, manipulados y almacenados de una manera estandarizada. Para todos ellos, los criterios de exclusión fueron los siguientes:

- Una significativa patología estructural cerebral en el examen *post mortem* realizado por un neuropatólogo cualificado o por técnicas de neuroimagen antes de la muerte del individuo

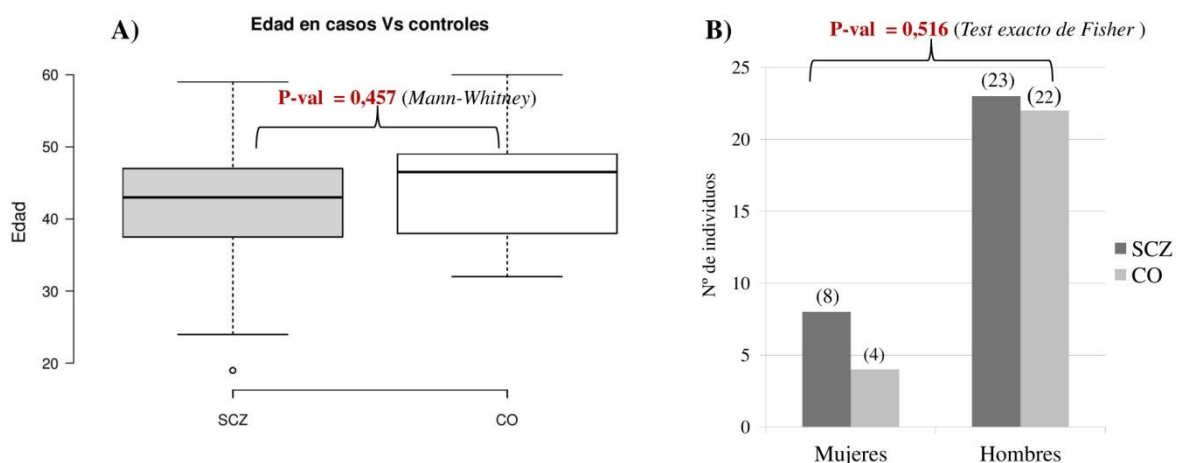
- Historial de deficiencias neurológicas focales antes de la muerte (esto es, algún problema en el funcionamiento del cerebro, la médula espinal y los nervios que afecta a un sitio específico, como el lado izquierdo de la cara, el brazo derecho o incluso un área pequeña como la lengua)
- Presencia de una enfermedad del sistema nervioso central (SNC) en la historia clínica de la que se podría esperar una alteración de la expresión génica de manera persistente
- Coeficiente intelectual (IQ) del individuo < 70
- Mala calidad de la muestra de ARN

En el caso de controles sanos, los criterios de exclusión adicionales fueron los siguientes:

- Edad menor de 30 años (y, por lo tanto, todavía en el período de máximo riesgo)
- Abuso de sustancias al menos 1 año antes de la muerte o presencia de cambios significativos relacionados con el consumo alcohol en el hígado

Los diagnósticos fueron hechos por dos psiquiatras experimentados, usando los criterios del DSM-IV y en base a los registros médicos y, en caso necesario, mediante entrevistas telefónicas adicionales con los miembros de la familia. Los diagnósticos de controles no afectados se basaron en entrevistas estructuradas por un psiquiatra de alto nivel con miembros de la familia para descartar diagnósticos psiquiátricos relacionados en el seno familiar.

Figura 3. 1. Comparación de edad y sexo en las muestras de casos (31) y controles (26) analizadas. Para la comparación de edades, se empleó un test no paramétrico de Mann-Whitney; y para la comparación de proporciones de hombres y mujeres en casos y controles, una prueba exacta de Fisher de dos colas.



Las características de las muestras utilizadas en este estudio se describen detalladamente en el **ANEXO I**. Dado que estas muestras se utilizaron para el análisis de expresión génica, se comprobó que no existiera sesgo alguno en el grupo de casos o en el de controles para alguna de las variables informadas que pudieran condicionar los resultados (**figura 3.1**). Los grupos de casos y controles mostraron edades similares ($42,3 \pm 8,83$ en casos frente a $44,6 \pm 7,09$ en controles; valor P en el test de Mann-Whitney = 0,457) y una proporción de hombres y mujeres comparable (74,1% en casos frente a 84,6% en controles; P en el test exacto de dos colas de Fisher = 0,516).

3.1.2. Preparación de las muestras para su utilización

Las muestras de ADN de los estudios I-III se obtuvieron de extracción de sangre total. Las muestras recopiladas de casos y controles durante los últimos años fueron recibidas como sangre fresca, congeladas, y almacenadas a -80°C . Posteriormente, se extrajo ADN mediante el robot de extracción *Chemagic Magnetic Separation Module I*, siguiendo las instrucciones para del protocolo del fabricante Chemagen (Baesweiler, Alemania). Una vez el ADN fue extraído, se cuantificó mediante el espectrofotómetro Nanodrop, obteniendo datos aproximados de la concentración y de la pureza del ADN, mediante análisis de la absorbancia.

Para los estudios I-II, se prepararon placas de muestras de casos y controles a $20\text{ ng}/\mu\text{L}$ de los individuos utilizados, a partir de los datos de concentración previamente obtenidos.

En el caso del estudio III, se prepararon 86 pools de casos y otros 86 pools de controles, que se amplificaron y secuenciaron conjuntamente. Por ello, para garantizar la equimolaridad de las muestras integrantes de los pools, se recuantificaron las 516 muestras de casos y las 516 de controles analizadas en este estudio mediante 3 medidas independientes utilizando el ensayo dsDNA BR del fluorímetro Qubit 2.0 (Invitrogen, Carlsbad, CA), siguiendo las instrucciones del fabricante. La media de las tres medidas se utilizó para elaborar una colección a $50\text{ ng}/\mu\text{L}$, y éstas se agruparon en pools de 6 muestras.

Las muestras utilizadas en el estudio IV fueron preparadas previamente a la realización del análisis de datos de secuenciación de las mismas, por lo que dicho proceso queda fuera de los métodos descritos en este trabajo de tesis.

3.2. SELECCIÓN DE REGIONES REGULADORAS

3.2.1. Elementos ultraconservados en genes asociados a esquizofrenia. *TCF4*

Se realizó una búsqueda de genes asociados a esquizofrenia presentes en la base de datos de genes afectados o flanqueados por UCEs (<https://users.soe.ucsc.edu/~jill/ultra.html>). El gen *TCF4* fue escogido para el estudio de variación rara en regiones ultraconservadas por su implicación en la esquizofrenia (Steinberg y cols., 2011a) y por albergar hasta dos elementos diferentes a lo largo de su estructura: UC435 (chr18:53,089,931-53,090,157; hg19), de 227 pb de longitud y que se encuentra en su totalidad en región no codificante intrónica y UC436 (chr18:53,254,241-53,254,450; hg19), de 210 pb de longitud con 92 pb en región exónica (UTR + CDS) y 118 pb en región intrónica. El elemento UC435 fue escogido para su estudio al ser éste enteramente no codificante y, en mayor medida, potencial regulador de la expresión génica.

Se realizó la identificación de secuencias homólogas de UC435 en los genomas de 22 mamíferos placentarios y 3 marsupiales adicionales comparando dichas secuencias mediante la herramienta *BLAST* del servidor del NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) con la secuencia de UC435 de la base de datos genómica de Refseq. Por otra parte, también se realizó la búsqueda de variantes en los datos de la Fase I del Proyecto de los 1000 genomas, que engloban 14 poblaciones procedentes de Europa, Asia Oriental, África subsahariana y América (1000 Genomes Project Consortium, 2012). Esta se llevó a cabo mediante la búsqueda a través de la base de datos del navegador del UCSC introduciendo las coordenadas de UC435.

3.2.2. Selección de variantes en elementos acelerados en humanos (HAEs) de genes asociados a esquizofrenia. *NPAS3*

Los elementos acelerados en humanos (HAEs), que se definen como regiones del genoma cuya secuencia se haya conservada en mamíferos pero que acumulan un número elevado de mutaciones en el linaje humano, han sido objeto de búsqueda y caracterización en diversos estudios (Bird y cols., 2007; Bush y cols., 2008; Lindblad-Toh y cols., 2011; Pollard y cols., 2006a; Prahbakar y cols., 2006).

El gen *NPAS3*, relacionado con la aparición de la esquizofrenia (Kamnasaran y cols., 2003; Pickard y cols., 2009), se escoge para el estudio de variación genética a lo largo de elementos HAEs, al ser éste el gen con mayor número de estos elementos (14) en todo el genoma (Kamm y cols., 2013a). Además, la mayoría de estas regiones funcionan como elementos reguladores de la expresión en pez cebra (Kamm y cols., 2013a) y ratón (Kamm y cols., 2013b).

Se realizó una búsqueda de SNV presentes en la base de datos de los 1000 genomas en poblaciones europeas (1000 Genomes Project Consortium, 2012) en los HAEs de *NPAS3*, de acuerdo con las posiciones genómicas en Kamm y cols., 2013. De este modo, un total de 27 variantes en población europea fueron seleccionadas de un total de 96 registradas en la base de datos para todas las poblaciones.

Además, por otra parte, los SNP en *NPAS3* con alguna evidencia de asociación con trastornos psiquiátricos en los GWAS también fueron elegidos mediante la búsqueda en la base de datos de *Ricopili* (<http://www.broadinstitute.org/mpg/ricopili/>) y el catálogo de GWAS (Welter y cols., 2014). De este modo, 6 SNP adicionales fueron también seleccionados para su estudio: rs1315132 ($P = 0,0049$ en trastorno por déficit de atención e hiperactividad (Neale y cols., 2010)), rs927327 ($P = 0,00038$ en esquizofrenia (Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011)), rs7155832 ($P = 2,27 \times 10^{-5}$ en trastorno bipolar (Sklar y cols., 2011)), rs8017055 ($P = 0,000851$ en depresión mayor (Ripke y cols., 2013b)), rs8015929 (con $P = 4 \times 10^{-6}$ en GWAS de trastorno bipolar (Ferreira y cols., 2008)) y rs4982029 ($P = 4 \times 10^{-6}$ en un estudio combinado de trastorno bipolar, depresión mayor y esquizofrenia (Huang y cols., 2010)).

En el caso de las variantes que presentan alto desequilibrio de ligamiento (DL) entre ellos ($r^2 \geq 0,70$), sólo una de ellas se retuvo para el diseño de los ensayos. De las 33 variantes, 3 fueron eliminadas por ello: rs8015929 (Ferreira y cols., 2008), rs185215180 (perteneciente al HAE HACNS658) y rs12891186 (perteneciente al HAE 2xHAR157) fueron eliminadas por estar en LD con rs4982029 ($r^2 = 1$, en todas las poblaciones europeas de los 1000 genomas), rs190286431 (presente en población italiana de Toscana (TSI) a $r^2 = 1$) y rs462723 (presente en población europea residente en Utah (CEU) a $r^2 = 0,74$), respectivamente. Finalmente, un total de 30 variantes (25 SNV en elementos HAEs de *NPAS3* en población europea de los 1000 genomas y 5 SNP de estudios de GWAS de trastornos psiquiátricos) fueron seleccionadas para el diseño de un estudio de genotipado.

3.2.3. Selección de genes de interés para la secuenciación de promotores mediante NGS

Los genes fueron seleccionados para la secuenciación de sus promotores basándose en dos fuentes diferentes: por un lado, aquellos que resultasen significativos en estudios de asociación de genoma completo o estudios similares con muestras de gran tamaño (Carrera y cols., 2012; O'Donovan y cols., 2008; Ripke y cols., 2013a; Steinberg y cols., 2011a) y, por otro, los presentes en CNV asociadas a esquizofrenia (Costain y cols., 2013; Degenhardt y cols., 2013; Guha y cols., 2013; Lee y cols., 2012b; Levinson y cols., 2011; Malhotra y cols., 2011;

Rees y cols., 2013; Vacic y cols., 2011). En lo relativo a estos últimos, dado que en la mayoría de las ocasiones la CNV afecta a la duplicación o delección de más de un gen, se seleccionaron para su estudio aquellos genes candidatos en base a los siguientes criterios

- Por un lado, se seleccionaron los genes con mayor probabilidad de estar involucrados en los fenotipos del sistema nervioso central de acuerdo con su función biológica relacionada con el mismo de las descritas en la base de datos *Gene Ontology* (Ashburner y cols., 2000), así como en la literatura.
- Se seleccionaron, preferentemente, aquellos genes cuya secuencia promotora fuese relativamente conservada (mediante los valores de conservación *PhastCons* (Siepel y cols., 2005) y *PhyloP* (Pollard y cols., 2010)).
- Por último, y como consecuencia de los múltiples diseños de PCR realizados, un factor determinante a la hora de la selección consistió en el mayor o menor rendimiento de las PCR en cada región, seleccionándose preferentemente aquellos genes candidatos cuyas regiones promotoras se amplificasen de forma más exitosa.

Las regiones promotoras de los genes seleccionados se basaron en los transcritos descritos en la base de datos de Gencode versión v14. No obstante, en genes de gran número de transcritos descritos según esta base de datos, se seleccionaron aquellos mejor descritos por estar también identificados en la base de datos de RefSeq.

El conjunto final de genes seleccionados consistió en los descritos a continuación:

- *ZNF804A*, *TCF4*, *SLC39A8*, *AKT3*, *DPYD*, y el *MIR137*, de estudios de GWAS en esquizofrenia (Carrera y cols., 2012; O'Donovan y cols., 2008; Ripke y cols., 2013a; Steinberg y cols., 2011a).
- *BCL9* en 1q21.1, *CYFIP1* en 15q11, *MAGEL2* y *CHRNA7* en 15q11.2-13.3, *KCTD13*, *MVP*, y *TAOK2* en 16p11.2, *PMP22* en 17p12, *AATF*, *HNF1B*, y *LHX1* en 17q12, *DGCR8* y *COMT* en 22q11.2, *DLG1* y *PAK2* en 3q29, 7q36 en *CNTNAP2*, y *NRXN1* en 2p16, de estudios de CNV asociadas a la enfermedad.

Los diseños de PCR para la correcta amplificación de los 30 amplicones de regiones promotoras de los 23 genes seleccionados, así como aquellos que no fueron finalmente incluidos, se describen en el **apartado 3.6**

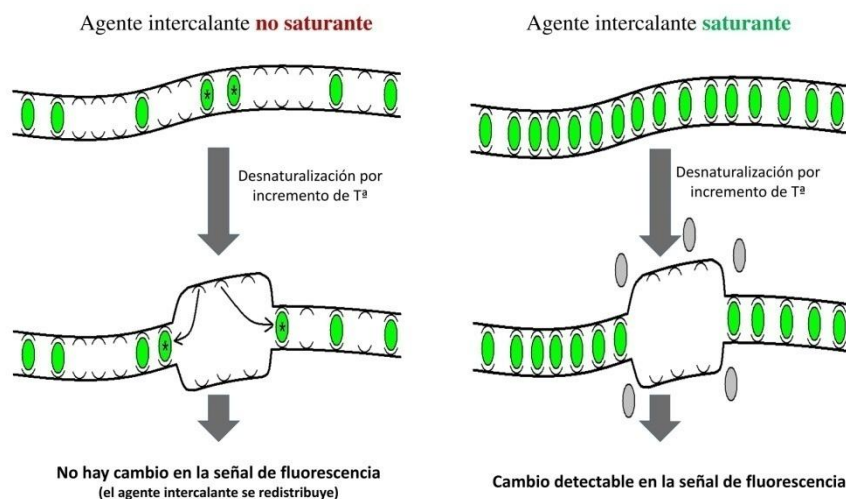
3.3. DETECCIÓN DE MUTACIONES MEDIANTE HRMA

3.3.1. Introducción teórica

HRMA (del inglés *High Resolution Melting Analysis*) es una técnica post-PCR para la identificación de variantes mediante la detección de cambios en el comportamiento de la disociación de ADN al aumentar la temperatura cuando existe alguna variante en su secuencia (Herrmann y cols. 2006, Vossen y cols. 2009, Wittwer 2009). Se basa en el principio de que la temperatura de desnaturalización de un amplicón determinado es dependiente de la composición de la secuencia del mismo (Ririe y cols., 1997).

Este método implica la realización previa de una amplificación mediante PCR en presencia de agentes intercalantes saturantes y fluorescentes que se unen entre la doble cadena de hebras hibridadas de ADN, como *LCGreen*. Al aumentar la temperatura necesaria para desnaturalizar la doble hebra de ADN, tiene lugar la pérdida del agente intercalante y, por consiguiente, una progresiva reducción en la señal de fluorescencia (**figura 3.2**). Este patrón de reducción en la señal luminiscente se monitoriza a lo largo del gradiente de temperatura aplicado (Reed y Wittwer, 2004).

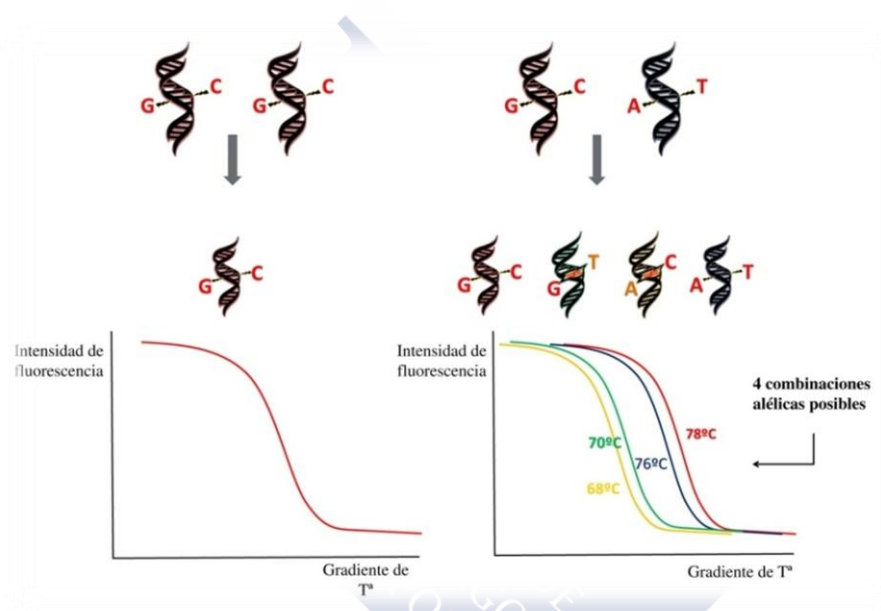
Figura 3. 2. Fundamento de la pérdida de fluorescencia al desnaturalizar la doble hebra de ADN.



El resultado es un perfil del patrón de desnaturalización que varía en función de la secuencia que se desee amplificar. En este sentido, un mayor contenido de GC a lo largo de la secuencia, y debido al mayor número de enlaces de hidrógeno entre las bases hidrogenadas de los nucleótidos, provoca que la desnaturalización de la doble hebra de ADN ocurra a

temperaturas más elevadas. En presencia de variantes a lo largo de la secuencia, tras el proceso de amplificación mediante PCR se formarán homodúplex y heterodúplex, cuando las secuencias hibriden correcta o incorrectamente en la posición de la variante, respectivamente (**figura 3.3**). Esto contrasta con la única posibilidad de homodúplex existente si no hay cambios en la secuencia. Por ello, el análisis de perfiles de desnaturalización permite la distinción de heterodúplex y homodúplex y la consiguiente detección de variantes en la secuencia de ADN (Graham y cols. 2005), convirtiendo a esta técnica en una alternativa muy económica para el barrido de mutaciones infrecuentes o raras en una secuencia de interés.

Figura 3. 3. Cambio en la curva de fluorescencia en presencia de variantes genéticas en la secuencia analizada.



Una de las principales ventajas de esta técnica es que no es destructiva, pudiendo ser recuperadas las muestras al finalizar el análisis. Sin embargo, para la implementación de la técnica de HRMA es necesario que se cumplan una serie de requisitos muy importantes:

- Buena calidad de la PCR: El producto de la reacción de amplificación debe ser único y específico, presentar un elevado rendimiento y en ausencia de dímeros.
- El ADN utilizado en las diferentes muestras analizadas debe tener una concentración homogénea y estar en buen estado.
- La optimización del proceso de PCR debe realizarse en presencia del agente intercalante, ya que este producto de amplificación es el que se analizará directamente en HRMA.

3.3.2. Amplificación del elemento UC435 de *TCF4* para el análisis de mutaciones mediante HRMA

El diseño de los *primers* de extensión para la amplificación de la región del elemento conservado UC435 se realizó mediante el programa Primer3web (Untergasser y cols., 2012), de libre acceso *online* (<http://primer3.ut.ee/>). Previamente, la secuencia del elemento UC435 se extrajo de la base de datos de los 481 UCEs del estudio de Bejerano y cols. (<https://users.soe.ucsc.edu/~jill/ultra.html>) y se convirtió de las coordenadas genómicas de Julio de 2003 (NCBI34/hg16) a las de Febrero de 2009 (NCBI37/hg19) mediante la herramienta *convert* del browser de UCSC (<https://genome.ucsc.edu/>). La secuencia de ADN de UC435 se extrajo mediante la herramienta *get DNA* del browser de UCSC, con 100 pb a mayores a ambos extremos 5' y 3' de la secuencia, y el total de 427 pb se utilizó para el diseño de cebadores en Primer3web. El programa fue utilizado con los parámetros de diseño por defecto, y las propiedades de los *primers* y localización de región diseñada (chr18:53089911-53090194) se describen en las **tablas 3.1 y 3.2**.

Tabla 3. 1. Secuencia y temperatura de fusión de los cebadores diseñados para la amplificación del elemento UC435.

Amplicón	Primer	Secuencia	Tª fusión
uc435	<i>Primer F</i>	CTGGATCTTGGAAGCCTGA	64,64
	<i>Primer R</i>	TCCTCATCATTCTGCCTGTAAA	63,8

Tabla 3. 2. Propiedades de la secuencia diseñada para la amplificación del elemento UC435. ¹ El elemento ultraconservado se extrajo de Bejerano y cols. (2004)

Gen	Elemento ¹	Coordenadas (hg19)	Longitud amplicón (pb)	% conservación (PhastCons46way)	% CG	Amplicón	Enzima utilizada
TCF4	uc435	chr18:53089911-53090194	284	100%	36,61	UC435	Go Taq (Promega)

Posteriormente se comprobó, mediante PCR *in silico* (<https://genome.ucsc.edu/cgi-bin/hgPcr>), que los *primers* diseñados no amplificaban en ninguna otra región del genoma que no fuese la deseada en TCF4. Además, mediante la herramienta *BLAT* del browser de UCSC (<https://genome.ucsc.edu/cgi-bin/hgBlat>) se comprobó que las secuencias de los *primers*, individualmente, no se hibridaban en ninguna otra región del genoma humano. Por último, se comprobó que no existían variantes conocidas a lo largo de las secuencias de los cebadores. Para ello se utilizó la herramienta *SNPCheck* (<http://ngl.org.uk/SNPCheck/snpcheck.htm>).

Las condiciones de reacción de PCR se optimizaron en un termociclador Applied Biosystems® Veriti® (Life technologies, California). En este proceso se utilizaron los reactivos

LightScanner *Master Mix* (Idaho Technologies, Salt Lake City, Utah), que incluye a la enzima polimerasa *Taq*, un anticuerpo para la activación de la enzima mediante su liberación (por la metodología *Hot Start*), dNTPs, y MgCl_2 10 mM. Junto con la mezcla *Master Mix*, se añadieron las disoluciones de los *primers* a 10 μM , el colorante saturante LC Green Plus 10X (Idaho Technologies, Salt Lake City, Utah) y 20 ng de ADN, en las cantidades establecidas según el protocolo establecido para la enzima empleada. Finalmente, se añade aceite mineral, para evitar la evaporación de los reactivos, a los 13 μL de mezcla de reacción (**tabla 3.3**). Las reacciones se llevaron a cabo en placas opacas de 96 pocillos (Sigma Aldrich®, St. Louis, Missouri, USA).

Tabla 3. 3. Condiciones de amplificación en PCR de la región deUC435 en TCF4. En la tabla se representan los valores de concentración inicial ($[I_i]$) y concentración final de reactivos ($[I_f]$) en μM , así como el volumen añadido de los mismos en μL .

<i>uc435</i>	$[I_i]/\mu\text{M}$	$[I_f]/\mu\text{M}$	$V/\mu\text{L}$
Master Mix	-	-	6
LC Green	-	-	1,2
PRIMER F	10	0,48	0,62
PRIMER R	10	0,48	0,62
DNA muestra (<i>ng/μL</i>)	20		1
H ₂ O mQ			3,56

Las condiciones de la reacción de PCR para la amplificación de UC435 se optimizaron a partir del protocolo de recomendado del manual de HRMA de *Light scanner* (Idaho Technologies Salt Lake City, Utah) mediante variaciones en las temperaturas de alineamiento, los tiempos de extensión y en el número de ciclos de reacción (**tabla 3.4**), chequeando los productos de la optimización en chips de *Agilent 2100 Bioanalyzer* (Agilent technologies Santa Clara, California). Este equipo, conectado a un PC, permite la separación de los productos de reacción mediante electroforesis de una forma rápida y precisa. La elevada resolución de la carrera electroforética y la estimación de la concentración a partir de la intensidad de cada banda son factores muy ventajosos a la hora de optimizar, mediante esta tecnología, la reacción de PCR con exactitud.

El protocolo optimizado se utiliza para la amplificación de un conjunto de 528 casos esquizofrénicos, en 6 placas, y 270 controles sanos, en 3 placas. Los productos de reacción se chequean mediante electroforesis en *Agilent 2100 Bioanalyzer* y se llevan al equipo de HRMA *light scanner*.

Tabla 3. 4. Condiciones de PCR optimizadas de la amplificación del elemento UC435 de TCF4 para HRMA.

Amplicón	Desnaturalización inicial	Ciclos (35x)			Extensión final	Final
		Desnaturalización	Alineamiento	Extensión		
UC435	94 °C	94 °C	59 °C	72 °C	72 °C	12 °C
	2 min.	30 seg.	30 seg.	1 min.	7 min.	2 min.

3.3.3. Análisis de variantes genéticas en UC435 mediante HRMA.

El perfil de fluorescencia esperado para la secuencia amplificada del elemento UC435 (chr18:53,089,911-53,090,194) se analizó mediante el software *uMELT* (Dwight y cols., 2011; <https://www.dna.utah.edu/umelt/um.php>).

La búsqueda de mutaciones en UC435 en 528 casos y 270 controles se realizó en el equipo de *HRMA LightScanner®* (Idaho Technologies, Salt Lake City, Utah), el cual es compatible con el termociclador *Veriti®* (Life technologies, California) previamente utilizado, mediante el software *Call -IT 2.0* (Idaho Technologies, Salt Lake City, Utah). El uso de esta técnica es muy indicado en el presente caso de la detección de variantes raras, pues existe un grupo definido de perfiles de fluorescencia similares (sin variante) que se diferenciaría claramente de las muestras con posibles variantes.

En el análisis, en primer lugar se marcan los negativos de reacción. Seguidamente, y tras la calibración del instrumento, se estudian los patrones de pérdida de fluorescencia al desprenderse el agente intercalante LC Green en un barrido de temperaturas desde 45 °C hasta 94 °C. Este proceso se realizó acuerdo con las instrucciones del fabricante (Idaho Technologies, Salt Lake City, Utah), donde cada muestra de 12 µL se desnaturizó a una velocidad de 0,1 °C/segundo adquiriéndose 14 valores de fluorescencia por °C incrementado de temperatura. Los perfiles de las curvas de desnaturalización de cada muestra se normalizan para que tengan misma temperatura inicial y final y, finalmente, se agrupan los distintos perfiles de desnaturalización (si hay diferencias) en grupos distinguibles que se hacen más visibles mediante la transformación de las curvas normalizadas de fluorescencia en curvas de gradiente de señal de fluorescencia entre las distintas muestras.

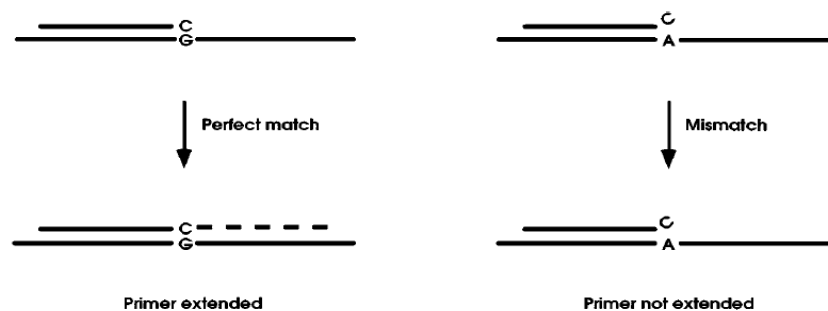
3.4. TECNOLOGÍA DE GENOTIPADO DE SEQUENOM

3.4.1. Introducción teórica

La tecnología de Sequenom (Sequenom, San Diego, CA) permite el genotipado de entre 1 y 40 variantes genéticas por reacción, como SNP, inserciones y deleciones, o mutaciones mediante la utilización de la plataforma de espectrometría de masas *MassARRAY*. La base química de esta técnica se fundamenta en los siguientes pasos consecutivos:

- **Amplificación del ADN mediante una *PCR multiplex* (Assay designer).** Este proceso puede ser repetido unas 30 - 40 veces, tras el cual tiene lugar la purificación mediante la enzima *SAP* (del inglés *Shrimp Alkaline Phosphatase*), que defosforila los dNTPs en el extremo 5' para que no puedan ser incorporados por la polimerasa en fases posteriores.
- **Minisecuenciación de una sola base o *SBE*** (del inglés *Single Base Extension*). El fundamento de la técnica de minisequenciación consiste en que la secuencia de ADN que contiene el SNP sirve como molde para la sonda, que se une al mismo sólo si ambas son complementarias en el sitio polimórfico (**figura 3.4**) y con la ayuda de la ADN polimerasa, se va extendiendo. Monitorizando la extensión del *primer* se consigue identificar el alelo presente en la muestra de ADN.

Figura 3. 4. Extensión específica de primer en la minisequenciación (SBE)(Kwok, 2001)



- Detección mediante el **sistema MALDI-TOF** (del inglés *Matrix Assisted Laser Desorption/ Ionization Time of Flight*) en el *Autoflex* (Bruker Daltonics GmbH, Bremen, Alemania). Mediante espectrometría de masas, se detectan los productos a partir de su masa molecular, pudiendo discriminarse un alelo u otro. Las bases empleadas se modifican para que exista una diferencia entre ellas superior al mínimo detectable de 15 Da (Chen y Sullivan, 2003).

El genotipado en Sequenom de los trabajos aquí presentados se realizó en la plataforma del Centro Nacional de Genotipado (CEGEN) de Santiago de Compostela.

3.4.2. Genotipado de variantes en elementos acelerados en humanos (HAEs) de NPAS3.

De las 30 variantes seleccionadas para el estudio de la variación genética en *NPAS3* que se utilizaron para el diseño de un ensayo multiplex mediante Sequenom MassArray (Sequenom, San Diego, CA), 4 fueron excluidas. Las variantes rs73270956 (HAE HAR202), rs138639467 (HAE HAR21), rs140756409 (HAE HAR202) y rs57977791 (HAE HAR189-HACNS489) fueron eliminadas al no poder ser incorporadas al ensayo multiplex en Sequenom.

MassArray IPLEX se utilizó para determinar el genotipo de las variantes seleccionadas en un ensayo múltiple de alto nivel, siguiendo las instrucciones del fabricante.

3.4.3. Confirmación mediante genotipado de variantes de secuenciación dirigida en regiones promotoras.

La plataforma de Sequenom MassArray (Sequenom, San Diego, CA) fue utilizada para analizar la presencia de falsos positivos en las variantes detectadas por la plataforma de secuenciación de SOLiD, siguiendo el procedimiento detallado en el **apartado 3.6**.

Para ello, se seleccionaron 106 SNV cuya predicción funcional en RegulomeDB fuese 2, de modo que, escogiendo el grupo de variantes de mayor predicción funcional, se analizase la tasa de falsos positivos en la selección de variantes más propensas a serlo. De los 106, un total de 89 entraron finalmente en un ensayo de 4 multiplexes (multiplex W1 con 30 SNV, multiplex W2 con 29 SNV, multiplex W3 con 26 SNV y multiplex W4 con 4 SNV).

Las variantes fueron genotipadas en las muestras potencialmente afectadas (se genotiparon todas aquellas pertenecientes a cualquier pool que las contuviese). *MassArray IPLEX* se utilizó para determinar el genotipo de las variantes seleccionadas en un ensayo múltiple de alto nivel, siguiendo las instrucciones del fabricante.

3.5. SECUENCIACIÓN DE SANGER

3.5.1. Amplificación del elemento UC435 de TCF4 para la secuenciación de Sanger de las muestras con posibles variantes.

Las condiciones de reacción se optimizaron en un termociclador Applied Biosystems® Veriti® 96 wells (Life technologies, California). En el proceso optimizado de amplificación se utilizaron como reactivos dNTPs 10 mM, MgCl₂ 25 mM, disolución de los *primers* anteriormente utilizados en el estudio de HRMA del UC435, a 10 µM, y el kit de Promega (Wisconsin, USA) Gotaq® Flexi ADN Polymerasa, en los volúmenes finales indicados en la **tabla 3.5**. Finalmente, se añaden 20 ng de ADN de las muestras estudiadas. Las condiciones de reacción utilizadas previa optimización se describen en la **tabla 3.6**.

Los productos de reacción se chequearon en chips de *Agilent 2100 Bioanalyzer* (Agilent technologies Santa Clara, California), para evaluar su presencia y especificidad.

Tabla 3. 5. Condiciones de amplificación en PCR de la región de UC435 en TCF4 para secuenciación. En la tabla se representan los valores de concentración inicial ($[]_i$) y concentración final de reactivos ($[]_f$) en µM, así como el volumen añadido de los mismos en µL.

<i>uc435</i>	$[]_i / \mu\text{M}$	$[]_f / \mu\text{M}$	$V / \mu\text{L}$
5X Gotaq® Flexi buffer	5	1	5
Gotaq® ADN pol(U/µL)	5	0,025	0,125
MgCl ₂ (mM)	25	2	2
dNTPs	10	0,2	0,5
PRIMER F	10	0,6	1,5
PRIMER R	10	0,6	1,5
DNA muestra (ng/µL)	20	1,6	2
H ₂ O mQ	-	-	12,375

Tabla 3.6. Condiciones de PCR optimizadas de la amplificación del elemento UC435 de TCF4 para secuenciación.

Amplicón	Desnaturalización inicial	Ciclos (35x)			Extensión final	Final
		Desnaturalización	Alineamiento	Extensión		
<i>UC435</i>	94 °C	94 °C	57 °C	72 °C	72 °C	4 °C
	2 min.	30 seg.	45 seg.	1 min.	7 min.	∞

3.5.2. Secuenciación de las muestras con posibles variantes en UC435

El producto de amplificación se purificó para, posteriormente, continuar con la reacción de secuenciación y el análisis de su secuencia. Para ello, se empleó el kit de ExoSAP-IT® (Affymetrix, Santa Clara, California). Este kit de purificación contiene *Exonucleasa I* y *Shrimp Alkaline Phosphatase*, y 0,85 µL del mismo se añaden sobre 2,15 µL de los productos de amplificación distribuidos sobre una placa de 96 pocillos. En la reacción se utilizó de nuevo un termociclador Applied Biosystems® *Veriti® 96 wells* (Life technologies, California), y en él se llevaron a cabo una incubación a 37 °C para la purificación propiamente dicha (eliminación de los nucleótidos sobrantes y degradación de los cebadores), y una posterior a 80 °C para inactivar a la enzima ExoSAP-IT® (**tabla 3.7**).

Tras la purificación, se procedió a la realización de la reacción de secuenciación. Para ello, se empleó el kit *ABI Prims® BygDye® terminator v 3.1 Cycle sequencing* (Thermo Fisher Scientific, Massachusetts), y la reacción tuvo lugar de nuevo un termociclador Applied Biosystems® *Veriti® 96 wells* (Life technologies, California). Las condiciones se describen en la **tabla 3.8**

Tabla 3. 7. Condiciones de la reacción de purificación de productos amplificados de UC435 mediante el kit ExoSAP-IT.

Amplicón	Incubación I	Incubación II	Final
<i>UC435 (amplificado)</i>	37 °C 15 min.	80 °C 15 min.	4 °C ∞

Tabla 3. 8. Condiciones de la reacción de secuenciación de los productos purificados de UC435 mediante el kit BygDye.

Amplicón	Paso I	Ciclos (25x)			Final
		Paso II	Paso III	Paso IV	
<i>UC435(purificado)</i>	96 °C 1 min.	96 °C 10 seg.	50 °C 5 seg.	60 °C 1 min 15 seg.	4 °C ∞

Una vez secuenciado, el producto de reacción se precipitó y su secuencia se leyó en el *3730xl ADN Analyzer* de Applied Biosystems (Life technologies, California). Este equipo genera archivos de secuencia ABI, que se importan al software *Seq Scape® Version 2.5* de Applied Biosystems (Life technologies, California) para poder leer dichas secuencias. Mediante comparación de las secuencias con las de referencia en hg19 para la región secuenciada de UC435 extraída mediante la herramienta *get DNA* del browser de UCSC, se

comprobó si existen variantes genéticas que justifiquen las diferencias en las curvas de desnaturalización en las muestras seleccionadas.

3.5.3. Secuenciación para la confirmación de las 30 regiones promotoras diseñadas

Una vez fueron optimizadas todas las PCRs, se amplificaron los 30 promotores finalmente incluidos en el estudio en muestras de ADN de controles sanos, y se llevaron a secuenciar al servicio de secuenciación de la USC, donde su secuencia se leyó en un *3730xl ADN Analyzer* de Applied Biosystems (Life technologies, California).

3.6. SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS). PLATAFORMA DE SOLID

3.6.1. Introducción teórica

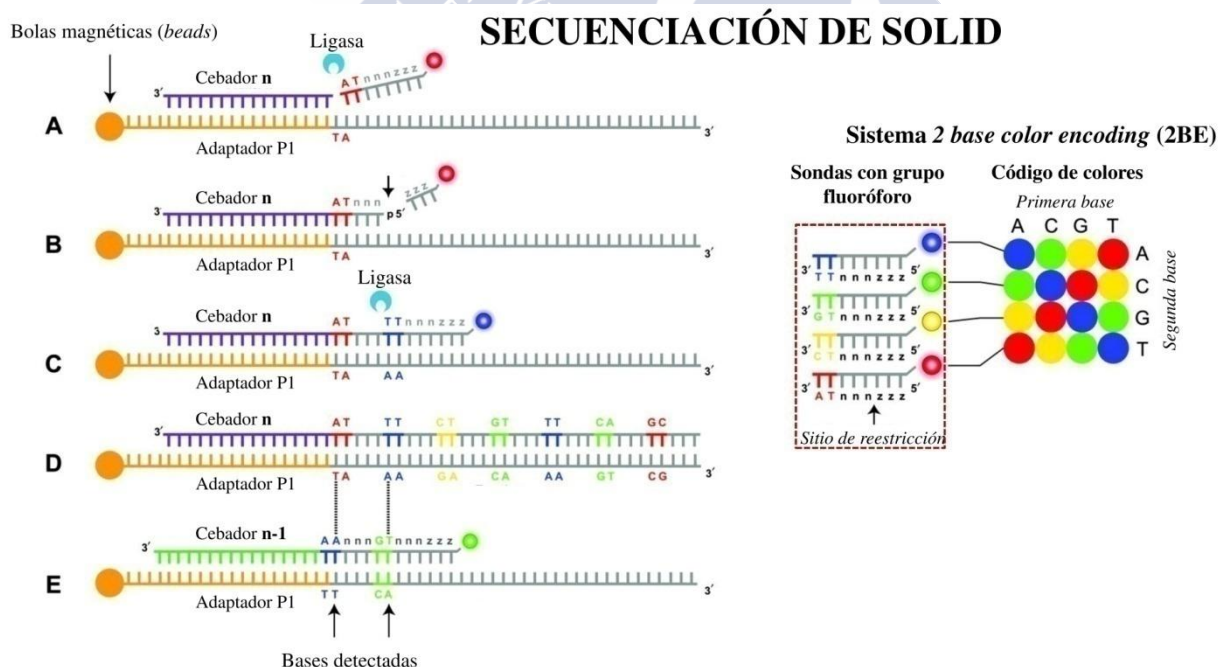
La plataforma secuenciación de nueva generación (NGS) SOLID (del inglés Sequencing by Oligo Ligation and Detection) salió al mercado en Octubre de 2007. Puede ser utilizada para secuenciación de genoma completo (WGS), resecuenciación de regiones concretas, estudios de expresión génica y del transcriptoma y estudios de inmunoprecipitación de la cromatina (ChIP). Cada carrera de secuenciación corrida en este instrumento requiere aproximadamente una semana y produce 3-4 Gb de datos de secuencia con una longitud de lectura de 25-35 pb, superando claramente el rendimiento de la tecnología de secuenciación de Sanger. En el proceso, tienen lugar las siguientes etapas (**figura 3.5**):

- En primer lugar, se realiza la **preparación de las librerías** de los fragmentos a amplificar. Esta etapa comienza con una sonicación inicial para fragmentar el ADN en fragmentos de 100-110 pb. A continuación, a estos fragmentos se les añaden dos adaptadores en ambos extremos (P1 y P2), pudiendo uno de ellos tener además un pequeño fragmento "código de barras" de forma que se pueda realizar la secuenciación multiplex. Se hace una selección de los fragmentos con adaptadores de tamaño adecuado mediante electroforesis en gel, para pasar a la siguiente fase, de amplificación de los mismos mediante PCR. Una vez finalizada la ePCR, se realiza una reacción de PCR cuantitativa (qPCR) para cuantificar la concentración de los productos amplificados.
- Tras la creación de las librerías, tiene lugar un proceso de **amplificación mediante PCR en emulsión (ePCR)**. De esta forma, los fragmentos de ADN ya preparados (con la unión de los adaptadores) se unen a bolas magnéticas *SOLiD™ P1 DNA Beads*, que

poseen uno de los dos primers (P1) unidos en su superficie para posibilitar la unión de los fragmentos que se van a amplificar. La reacción utiliza reactivos similares a los de una PCR normal (cebadores, enzima,...), además de una fase orgánica donde se integran, con agitación, esferas de fase acuosa que contiene todos los elementos anteriormente descritos y que actúan como reactores en paralelo para la amplificación. De este modo, los reactores son monoclonales o policlonales en función de si las bolas magnéticas incorporan cebadores exclusivos para un fragmento o distintos cebadores para amplificar varios fragmentos diferentes. Una vez finalizada la ePCR, se cuantifican los productos amplificados y tiene lugar la fase de enriquecimiento, mediante la cual se realiza un lavado y una subsiguiente centrifugación con glicerol al 60% para que las bolas sin producto amplificado se separen de las que sí lo tienen.

- En tercer lugar, tiene lugar la adición de un grupo químico en los extremos 3' de los fragmentos amplificados para la **deposición de las bolas amplificadas sobre las placas o slides** de la plataforma de SOLiD. Durante este proceso, se puede decidir la división del *slide*, y en vez de uno solo, emplear 4 o incluso 8 secciones distintas. La utilización de mayores densidades de bolas por *slide* resulta en un mayor nivel de rendimiento para el mismo sistema.

Figura 3. 5. Esquema de la secuenciación de SOLiD. Figura modificada de Voelkerding y cols. (2009)



- Finalmente, tiene lugar la **secuenciación por ligación**. Este procedimiento es el responsable de la baja tasa de falsos positivos de la plataforma de SOLiD en

comparación con otras coetáneas. Para ello, SOLID emplea el sistema 2BE (de inglés *2base color encoding*) de codificación de las combinaciones de dinucleótidos mediante colores. de esta forma se utilizan pequeños segmentos o sondas con todas las combinaciones de dinucleótidos posibles y constituidos de la siguiente forma: el par de nucleótidos en 5', seguido de 3 bases degeneradas y una cola de varias bases que comienza con un sitio de rotura enzimática y en el extremo 3' presenta un grupo fluoróforo correspondiente al color del dinucleótido en 5'. Tras la unión del cebador al fragmento a secuenciar, los distintos segmentos se unen en función de la secuencia de las dos primeras bases. Después de la unión del primer segmento, se rompe el mismo por el sitio de rotura y se registra la señal fluorescente para la secuencia de esos dos primeros nucleótidos. A continuación, el siguiente segmento se une a la posición del primero + 3 (correspondiente a las 3 bases degeneradas) y una ligasa une ambos, ocurriendo nuevamente la rotura del segundo segmento y la liberación y registro de la señal fluorescente. Este proceso se repite a lo largo de toda la secuencia a amplificar. Además, se utilizan cinco combinaciones distintas de *primers* que difieren en la longitud del mismo en el extremo 3' (primer n, primer n-1,..., primer n-4), lo que permite combinar la información de las 5 posibles secuencias "dinucleótido + 3 bases degeneradas" e inferir toda la secuencia del fragmento, además de servir de corrección de posibles errores en la secuenciación de alguna de ellas. La última versión de SOLID (SOLID 5500xl) incluye además otra secuenciación más, desde uno de los *primers* anteriores pero con un código de colores basado en combinaciones de 3 nucleótidos situados en la posición 1, 2 y 4 de cada segmento unido.

3.6.2. Secuenciación de genes de estudios GWAS y de CNV en esquizofrenia para el análisis de variantes reguladoras

3.6.2.1. Diseño de las regiones diana de secuenciación (targeted resequencing)

En el presente estudio, dado que se trató de una resecuenciación dirigida a regiones diana de interés, el enriquecimiento en las mismas se realizó mediante amplificación de PCR, con anterioridad a la creación de librerías.

Con el propósito de asignar de forma inequívoca las regiones reguladoras de un gen específico, se consideraron sólo aquellas en los sitios de inicio de la transcripción (TSS) de los genes seleccionados, que se corresponden a las secuencias promotoras de los mismos. Estas secuencias son, además, las regiones que explican en mayor medida la variabilidad en la

expresión de un gen (Ecker y cols., 2012), por lo que representan la aproximación ideal para el estudio de la variación genética como factor regulador de la expresión génica. Uno de los hallazgos más importantes del proyecto Encode fue determinar que estas secuencias promotoras reguladoras rodean el TSS, distribuyéndose simétricamente alrededor del mismo, y sin un sesgo hacia las regiones *upstream* que preceden al inicio del gen, tal y como se creía habitualmente (Birney y cols., 2007). Las señales de intensidad de los ensayos de inmunoprecipitación de la cromatina revelan un enriquecimiento que se extiende al menos 1000 pb en ambos sentidos, en los genes.

Por estas razones, se optimizaron los diseños de amplificación mediante PCR de las secuencias promotoras de los genes seleccionados para su estudio debido a evidencias en estudios de CNV y de GWAS abarcando, en cualquier caso, un mínimo de 2000 pb centrados en los TSS de los transcritos estudiados. El diseño de PCR en regiones TSS es de mayor dificultad que en las regiones codificantes debido a las características particulares de la secuencia en las primeras, tales como motivos repetitivos o un alto contenido de GC. Debido a eso, y dado que se procuró la amplificación de largas regiones de ADN en los elementos promotores (> 2000 pb) mediante PCR de largo alcance, se utilizaron ADN polimerasas HF (de alta fidelidad), de considerable menor ratio de errores (Cline y cols., 1996; Kitabayashi y cols., 2002; McInerney y cols., 2014), en combinación con otros reactivos como DMSO y buffers específicos para tales situaciones provistos en los reactivos de cada una de las enzimas utilizadas. En concreto, fueron tres las enzimas probadas en los transcritos amplificados: *PfuUltra II Fusion* (Agilent technologies, Santa Clara, CA), *Phusion Hot Start II* (New England Biolabs, Ipswich, MA) y *Q5 HF DNA polymerase* (New England Biolabs, Ipswich, MA).

El diseño de los cebadores de extensión se realizó mediante el programa Primer3web (<http://primer3.ut.ee/>). Previamente, se extrajo la secuencia de 2000 pb de cada una de las regiones promotoras de interés en coordenadas genómicas Febrero de 2009 (NCBI37/hg19) mediante la herramienta *get DNA* del navegador del UCSC, con 100 pb a mayores a ambos extremos 5' y 3' de la secuencia, y el total de secuencia seleccionada se llevó al diseño de cebadores mediante Primer3web, con los parámetros de diseño por defecto. No obstante, para el diseño de las regiones amplificadas y, por ende, de los cebadores de extensión, se tuvieron en cuenta las siguientes situaciones:

- En caso de solapamiento entre las regiones promotoras de más de un transcrito de un mismo gen (en las regiones de 2000 pb), se diseñó la amplificación para la región que englobase a todos ellos
- Se tuvo en cuenta la presencia de regiones repetitivas (que se extrajeron como 'N's en la secuencia de ADN de la región de interés), de forma que los *primers* de extensión no

hibridasen en estas regiones. En caso de que las regiones adicionales de 100 pb en donde se ubican los cebadores en 5' y 3' estuviesen compuestos por regiones repetitivas, se aumentó la longitud del amplicón diseñado hasta encontrar 100 pb en las que hubiera secuencia no repetitiva para el diseño del primer.

Una vez diseñada la pareja de *primers* en Primer3web, se comprobó mediante PCR *in silico* (<https://genome.ucsc.edu/cgi-bin/hgPcr>) que los *primers* diseñados no amplificaban en ninguna otra región del genoma que no fuese la deseada en los genes de interés. Además, mediante la herramienta *BLAT* del browser de UCSC (<https://genome.ucsc.edu/cgi-bin/hgBlat>) se comprobó que las secuencias de los *primers*, individualmente, no hibridaban en ninguna otra región del genoma humano.

Posteriormente, con el diseño del amplicón finalizado, se trató de optimizar el proceso de amplificación de las secuencias promotoras escogidas mediante la enzima *Phusion Hot Start II*. En el diseño de PCR, en los casos problemáticos (ausencia de amplificación o amplificación inespecífica), se siguieron las siguientes estrategias de optimización:

- Cambiando el buffer provisto por la casa comercial (Buffer HF, utilizado en primer lugar, se cambió por el Buffer GC para las regiones que no amplificaban).
- Añadiendo reactivos que facilitasen la apertura de regiones de posibles estructuras secundarias rígidas (mediante adición de DMSO en las regiones que no amplificaban).
- Modificando las condiciones de PCR: tiempos de extensión, temperaturas de alineamiento, tiempo de desnaturalización inicial.
- Modificando el tipo de PCR empleado: PCR anidada, una variante de la PCR convencional que comprende dos rondas de amplificación con distintos pares de cebadores en cada una, con el fin de incrementar la sensibilidad y la especificidad de la detección. Esta variante fue empleada en aquellos transcritos con amplificación inespecífica en cualquiera de las condiciones anteriores.

En los casos en los que ninguna de las anteriores opciones permitió una amplificación de las regiones seleccionadas específica, reproducible y con buen rendimiento, se modificó el diseño mediante las siguientes alternativas, combinándose a su vez con las 4 opciones destacadas anteriormente cuando correspondiese:

- Cambiando la enzima utilizada. La enzima *Phusion Hot Start II* se modificó por la *Q5 HF DNA polymerase*, en las condiciones básicas recomendadas o mediante las pautas anteriormente descritas para los casos problemáticos. En aquellos casos en donde tampoco esta nueva enzima arrojó resultados más favorables, se probó la tercera opción,

PfuUltra II Fusion, probándose igualmente los cambios anteriormente descritos en aquellos donde persistía el problema de amplificación

- Cambiando la pareja de cebadores. En los casos de amplificación inespecífica, se probó a utilizar cebadores de mayor longitud, cambiando la longitud por defecto de 18-20 a 22-24 pb en el *software*Primer3web. Por la contra, en casos en donde no hubiera habido amplificación de ningún tipo, se escogió la siguiente pareja de cebadores de la lista de resultados iniciales del primer3 para ese transcrito.
- Reduciendo la longitud de los amplicones. En los casos en los que se hubiesen juntado varias regiones de un gen al solaparse los promotores de varios transcritos del mismo, se probó a separar las regiones promotoras en unidades de menor longitud, tratando de cubrir al máximo la región pero sin llegar a solaparse regiones en distintos diseños de amplicones, de forma que ninguna región estuviese sobrerrepresentada en el conjunto final de promotores.

Las múltiples opciones probadas para cada uno de los genes deseados concluyeron finalmente en la selección de las 30 regiones promotoras de mejor rendimiento en la reacción de PCR del total de amplificaciones probadas. Estas regiones finales, se correspondieron con las regiones promotoras de 23 genes relacionados con la etiología de la esquizofrenia a partir de estudios GWAS y de CNV, comprendiendo un total de 77854 pb de secuencia reguladora. Dado el gran número de PCRs implementadas, las condiciones de las mismas para cada una de las regiones finalmente amplificadas, así como los volúmenes de reactivos empleados se describen en el **ANEXO III - IV**.

A continuación se describen las regiones escogidas para su amplificación en el presente trabajo de secuenciación de promotores por obtener los mejores rendimientos (**tabla 3.7**), al igual que aquellas que no se incluyeron finalmente (**tabla 3.8**), de la selección de genes procedentes de CNV y de estudios de GWAS en esquizofrenia.

Las parejas de primers diseñadas para la amplificación de las 30 regiones reguladoras estudiadas se describen en la **tabla 3.9**.

Tabla 3. 9. Propiedades de las regiones promotoras seleccionadas y finalmente amplificadas en el presente estudio

Gen	CNV/GWAS	Coordenadas (hg19)	Longitud amplicón (pb)	% conservación (PhastCons46way)	% CG	Amplicón	Enzima utilizada
<i>BCL9</i>	1q21.1	chr1:147011802 - 147014503	2702	58,02	49,48	AMP6	PHUSION (GC)
<i>PAK2</i>	3q29	chr3:196465367 - 196468056	2690	12,5	52,64	AMP19	pfu Ultra II
<i>DLG1</i>	3q29	chr3:196909183 - 196912067	2885	34,02	33,66	AMP21	PHUSION
<i>CNTNAP2</i>	7q36.1	chr7:145811906 - 145814893	2988	21,27	50,33	AMP29	pfu Ultra II
<i>CYFIP1</i>	15q11	chr15:22954356 - 22957235	2880	22,27	55,59	AMP31	Q5
<i>MAGEL2</i>	15q11.2-13.1	chr15:23889737 - 23892758	3022	29,83	67,07	AMP32	pfu Ultra II
<i>GABRA5</i>	15q11.2-13.1	chr15:27109972 - 27113828	3857	10,02	55,43	AMP39	PHUSION
<i>CHRNA7</i>	15q13.3	chr15:32403080 - 32405906	2827	6,4	45,42	AMP8	PHUSION
<i>MVP</i>	16p11.2	chr16: 29830829 - 29833231	2403	9,61	54,02	AMP12	PHUSION
<i>KCTD13</i>	16p11.2	chr16: 29935927 - 29939731	3805	8,85	52,1	AMP5bis	Q5
<i>TAOK2</i>	16p11.2	chr16:29983649 - 29986965	3317	37,44	55,5	AMP9bis	Q5
<i>TAOK2</i>	16p11.2	chr16:29990266 - 29993049	2784	9,76	47,2	AMP10	PHUSION
<i>PMP22</i>	17p12	chr17:15164514 - 15167161	2648	10,37	57,48	AMP54	PHUSION
<i>LHX1</i>	17q12	chr17:35292670 - 35295481	2812	46,46	62,52	AMP23	Q5
<i>AATF</i>	17q12	chr17:35304687 - 35307633	2947	15,95	53,17	AMP24	pfu Ultra II
<i>HNF1B</i>	17q12	chr17:36103597 - 36106638	3042	25,25	60,82	AMP25	pfu Ultra II
<i>COMT</i>	22q11.21	chr22:19948515 - 19951552	3038	9,36	62,05	AMP42	PHUSION
<i>DGCR8</i>	22q11.21	chr22:20067272 - 20068879	1608	12,59	66,04	AMP43	Q5
<i>AKT3</i>	GWAS	chr1:244006014 - 244007329	1316	40,36	39,36	AMP47B	PHUSION
<i>DPYD</i>	GWAS	chr1:98386056 - 98387257	1202	7,49	59,4	AMP45B	Q5
<i>MIR137</i>	GWAS	chr1:98510976 - 98512260	1285	60,44	60,31	AMP44B	pfu Ultra II
<i>TCF4</i>	GWAS	chr18:52969110 - 52971347	2238	22,92	41,96	AMP50X	PHUSION
<i>TCF4</i>	GWAS	chr18:52988640 - 52990185	1546	69,98	48,84	AMP49B	Q5
<i>TCF4</i>	GWAS	chr18:53176723 - 53179495	2773	24,43	34,73	AMP48	PHUSION
<i>TCF4</i>	GWAS	chr18:53256389 - 53257786	1398	61,94	46,35	AMP1A	PHUSION
<i>ZNF804A</i>	GWAS	chr2:185461601 - 185464534	2934	27,17	52,6	AMP2	Q5
<i>SLC39A8</i>	GWAS	chr4:103265511 - 103267169	1659	18,64	60,46	AMP52B	Q5
<i>NRXN1</i>	NRXN1	chr2: 50573210 - 50576396	3187	49,7	53,2	AMP17bis	Q5
<i>NRXN1</i>	NRXN1	chr2:51253993 - 51256786	2794	39,72	52,36	AMP3A	pfu Ultra II
<i>NRXN1</i>	NRXN1	chr2:51257789 - 51261055	3267	7,47	44,84	AMP3B	PHUSION

Tabla 3. 10. Propiedades de las regiones promotoras finalmente no incluidas en el presente estudio

Gen	CNV/GWAS	Coordenadas (hg19)	Longitud amplicón (pb)	% conservación (PhastCons46way)	% CG	Amplicón
<i>GABRB3</i>	15q11.2-13.1	chr15:27016846 - 27020230	3385	23.79	61.24%	AMP36
<i>GABRB3</i>	15q11.2-13.1	chr15:26960287 - 26963291	3005	0.4	46.49%	AMP37
<i>GABRB3</i>	15q11.2-13.1	chr15:26873754 - 26875742	1989	0.6	54.75%	AMP38
<i>GABRG3</i>	15q11.2-13.1	chr15:27215098 - 27218063	2966	4.02	56.98%	AMP40
<i>UBE3A</i>	15q11.2-13.1	chr15:25682252 - 25685451	3200	16.63	54.22%	AMP34
<i>CHRNA7</i>	15q13.3	chr15:32321203 - 32324231	3021	10.59	60.61%	AMP7
<i>DOC2A</i>	16p11.2	chr16:30020415 - 30024387	2983	19.9	68.19%	AMP22
<i>MAPK3</i>	16p11.2	chr16:30133792 - 30135056	1265	24.98	67.27%	AMP11E
<i>MAPK3</i>	16p11.2	chr16:30132915 - 30135999	3085	26.65	60.19%	AMP11D
<i>COMT</i>	22q11.21	chr22:19928223 - 19930140	1918	1.42	62.67%	AMP41
<i>MERKT</i>	2q13	chr2:112655042 - 112657566	2525	1.81	62.26%	AMP27
<i>DLG1</i>	3q29	chr3:197022429 - 197027142	4714	18.43	52.10%	AMP20
<i>FBXO45</i>	3q29	chr3:196295062 - 196296392	1331	63.94	67.69%	AMP18
<i>NRXN1</i>	NRXN1	chr2:50721743 - 50724670	2928	11.2	38.80%	AMP13
<i>CSMD1</i>	GWAS	chr8:4850606 - 4853839	3234	18.48	55.50%	AMP55
<i>ZEB2</i>	GWAS	chr2:145277396 - 145278715	1320	85.51	45.68%	AMP46
<i>CYFIP1</i>	15q11	chr15:22891729 - 22894116	2388	4.58	64.45%	AMP30
<i>CSMD2</i>	GWAS	chr1:34630147 - 34631433	1287	75.72	44.84%	AMP56B
<i>RB1CC1</i>	8q11.23	chr8:53626281 - 53627556	1276	25.64	33.27%	AMP51
<i>CNTNAP2</i>	7q36.1	chr7:147829138 - 147832185	3048	11.36	37.17%	AMP61
<i>BCL2L11</i>	2q13	chr2:111877003 - 111879930	2928	7.8	47.54%	AMP26
<i>PMP22</i>	17p12	chr17:15167142 - 15169976	2835	4.76	45.50%	AMP53
<i>NDN</i>	15q11.2-13.1	chr15:23930894 - 23933784	2891	32.11	52.85%	AMP33
<i>DOC2A</i>	16p11.2	chr16:30030982 - 30034336	3355	2.8	55.64%	AMP57
<i>KCTD13</i>	16p11.2	chr16:29932907 - 29935943	3037	7.16	45.74%	AMP4
<i>PRRT2</i>	16p11.2	chr16:29821747 - 29824731	2985	47.3	62.58%	AMP58
<i>ANAPC1</i>	2q13	chr2:112357308 - 112358642	1335	5.96	42.17%	AMP60

Tabla 3. 11. Secuencia y temperatura de fusión de los cebadores diseñados para la amplificación de todos los promotores estudiados

Amplicón	Primer	Secuencia	Tª fusión
AMP1A	<i>Primer F</i>	AAATGCATGCACACACTTTTTC	64,1
	<i>Primer R</i>	TAAAGTGGGCGTTATTTGCTTT	63,6
AMP2	<i>Primer F</i>	CAGACTCCCCCAAGTGGATA	64,28
	<i>Primer R</i>	TCAGGAGTGCTCTTCCGACT	64,6
AMP3A	<i>Primer F</i>	TCCAAAAAGCCTCCTGCTAA	62
	<i>Primer R</i>	CAGCTTCCTGCCAACAGTTT	64
AMP3B	<i>Primer F</i>	ACATTCCACACCATCAGCAA	64,6
	<i>Primer R</i>	AAACAGAACACAGAGTGGGTTG	63,4
AMP5bis	<i>Primer F</i>	TCTGATCTCCCACTTCCCTAAA	64,1
	<i>Primer R</i>	CACCTCAGCGACATAAGCAG	64
AMP6	<i>Primer F</i>	GCCTAGGCAGAGGAGACCTT	64,12
	<i>Primer R</i>	CTCCACCCCAATTCTCTGTA	64,28
AMP8	<i>Primer F</i>	CTGGATCTTGGAAGCCTGA	64,64
	<i>Primer R</i>	TCCTCATCATTCTGCCTGTAAA	63,8
AMP9bis	<i>Primer F</i>	CATGCACCTTCTGTCTCCA	64,4
	<i>Primer R</i>	AATCCTTTGGGGGATCATTC	64,1
AMP10	<i>Primer F</i>	TCGGGTGATGTTCTTCCTC	64,42
	<i>Primer R</i>	TTCGTCTGTGCAATGTGGT	64,68
AMP12	<i>Primer F</i>	TACACCACCTCCATCCCTTC	64,18
	<i>Primer R</i>	TTTTTGAGGCCTCACTTCT	64,17
AMP17bis	<i>Primer F</i>	TAAAGGGGATGACTCGGTTG	63,7
	<i>Primer R</i>	GTTGGGAGACACGGCATATT	63,7
AMP19	<i>Primer F</i>	TCCTGTACCCACTGCATCAA	64,2
	<i>Primer R</i>	TGCACAACCTTGATCCAAAA	64,1
AMP21	<i>Primer F</i>	AAATCAATCGTTGCCCATTC	63,5
	<i>Primer R</i>	GCTGAGGTTGGGACTTACGA	64,2
AMP23	<i>Primer F</i>	TTTCAGCTTCCGTCTGGTCT	63,9
	<i>Primer R</i>	GTCTGATGAGGCCAGAGAG	64
AMP24	<i>Primer F</i>	TTAAATGTGGGGTGGGAGAA	64
	<i>Primer R</i>	ACCTTTGTCCGTGTCCAAG	63,9
AMP25	<i>Primer F</i>	TCTGTCTTAAAGCGGAGTGG	64,3
	<i>Primer R</i>	CTCCAACCCAGCACTGTAT	63,9
AMP29	<i>Primer F</i>	AATTCAAACAGCAAAGGGACAT	63,8
	<i>Primer R</i>	TGCACTTCTAAAGGAATGCAGA	64,0
AMP31	<i>Primer F</i>	GAAGAAGAACGTCATCCAGAGG	64,3
	<i>Primer R</i>	ACTCAGGGAGGTCTGTCTAACG	63,9
AMP32	<i>Primer F</i>	GAGGATGACTTTCACCATCTCC	64,1
	<i>Primer R</i>	GGATCCACCTCCAATTGACTT	64,3
AMP39	<i>Primer F</i>	CAGTCTGCATTACAGAAGACC	64,2
	<i>Primer R</i>	GTGATAAAGCCAAAAGCTCCAG	64,1

AMP42	Primer F	CACTAGTAGGGATGTCCCTTGG	63,9
	Primer R	TTTCTTTAGGGTTCTGGGATGA	63,8
AMP43	Primer F	CCTGTATTAGCAAAGGGGACAG	63,8
	Primer R	AGAAAACCCGTACACTTGGACA	64,8
AMP44B	Primer F	CGAAAAGAAAATATGCGTTCCT	63,3
	Primer R	TAAGCCTCACTTCTCCCAAAG	63,7
AMP45B	Primer F	TACAGGAGGTGAAGGGGTAGAA	64
	Primer R	GCAAAAACGAAGAGCAACTTTT	63,7
AMP47B	Primer F	TTCCCCTACAACCATTCAAACCT	64,2
	Primer R	TCTTGGAACAGGCAGAGTAAGA	64,1
AMP48	Primer F	GCAGCCCCCAGAATTATAAAC	63,6
	Primer R	TTTGCAACTGTTGACCACATC	64,0
AMP49B	Primer F	TTTCGCTCTAGGATCCACTCTC	64
	Primer R	AAAATGTGCATGAGATGTTTGC	64,1
AMP50X	Primer F	ATGAGCTGATTCATCCTCGTTT	62,1
	Primer R	CCTGCTGCTTTGTTGCTCTA	61,4
AMP52B	Primer F	GCACTTTTCCCAAATAAAAAACG	63,5
	Primer R	CAAACCTGTTGACATCACACCT	64,2
AMP54	Primer F	TAGGGGTGAGGGGAGTAGGT	64
	Primer R	CTTGGCATATAACGGGCACT	64

3.6.2.2. Amplificación de las regiones diana de secuenciación (targeted resequencing)

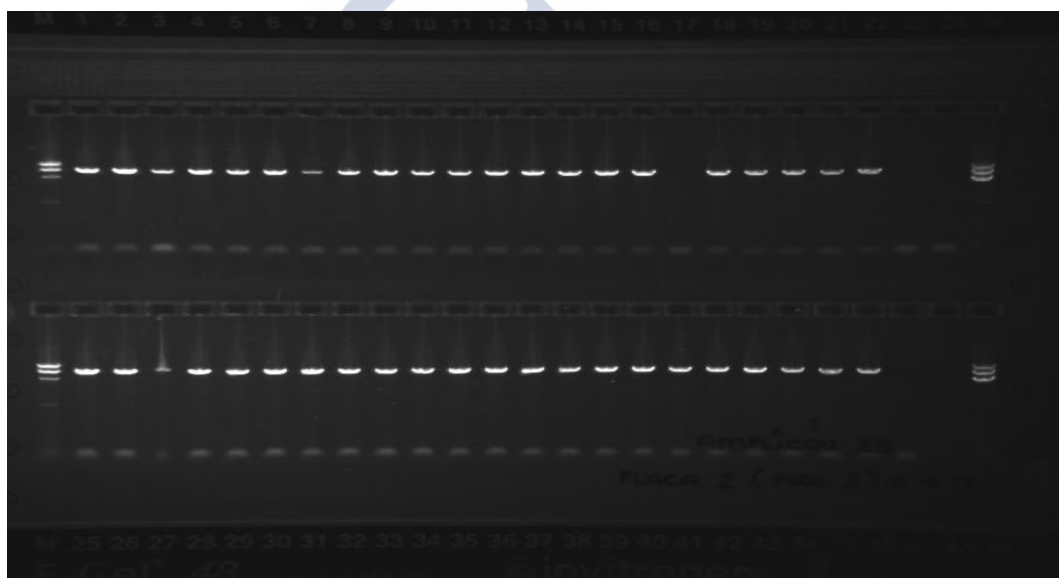
Una vez optimizadas las reacciones de PCR que se llevarían a cabo en el presente estudio, se realizó la amplificación de las regiones a estudiar en la población de 516 casos esquizofrénicos y 516 controles, agrupados en pools de 6 muestras. Con anterioridad a la amplificación final en los 172 pools, se comprobó mediante secuenciación de Sanger que las secuencias amplificadas se correspondían con las regiones de interés (**apartado 3.5.3**). Una vez hecho esto, los 30 amplicones optimizados en 86 pools de casos y otros tantos de controles se amplificaron, organizados en 2 placas de 96 pocillos y de forma intercalada para evitar posibles sesgos en ambos grupos en función de la posición de las muestras en las placas. Las últimas columnas de cada placa estuvieron compuestas por controles negativos, y en el interior de ambas placas se incluyeron dos blancos de PCR.

Las amplificaciones fueron chequeadas en geles de agarosa prefabricados E-gel (Invitrogen, Carlsbad, CA) del 1% o del 0,8%, tal y como se ejemplifica en la **figura 3.6**. Aquellas muestras con rendimiento inferior al esperado, se repitieron antes de dar por finalizado el proceso de amplificación.

Las concentraciones de los productos de amplificación, en las condiciones optimizadas, se midieron para cada uno de los 30 amplicones mediante ensayo de fluorimetría utilizando Qubit 2.0 (Invitrogen, Carlsbad, CA). De este modo, el volumen necesario para la preparación final de 2 placas que contuviesen, en una cantidad equimolar, la totalidad de las regiones amplificadas en los 172 pools de muestras. De este modo, se construyeron dos placas con unas concentraciones 305,6 pM.

Una vez construidas ambas placas finales, se secuenciaron siguiendo los protocolos de purificación y creación de librerías establecidos en la metodología correspondiente.

Figura 3. 6. Chequeo mediante E-gel de la amplificación de AMP23 para 44 pools (media placa de las 2 amplificadas en cada uno de los amplicones). En ambos lados se cargaron marcadores de tamaño y concentración. El pocillo 17 de la fila superior corresponde con el blanco de PCR presente en media placa.



3.6.2.3. Secuenciación de las regiones promotoras (targeted resequencing)

Las 30 regiones promotoras amplificadas en 172 pools de seis muestras (516 casos y 516 controles), y agrupadas en 2 placas con cantidades equimolares de cada región estudiada, se resecuenciaron mediante la plataforma SOLiD 5500xl (Applied Biosystems, Carlsbad, CA), empleando una metodología paired-end con reads de 75 y 25 pb, y basada en la química de llamadas exactas (ECC, del inglés *Exact Call Chemistry*), siguiendo las instrucciones del fabricante. Para ello, previamente, se procedió a la purificación de los productos de PCR, creación de librerías y seguimiento de los pasos descritos en la introducción teórica de este apartado para la plataforma SOLiD (**apartado 3.6.1**).

Las dos placas con los promotores amplificados en los 1032 individuos se purificaron mediante el kit de purificación de PCR QIAquick (QIAGEN, Venlo, Netherlands). Este kit de purificación contiene columnas de centrifugación, tampones, y tubos de recogida para la purificación a base de membranas porosas de los productos de PCR de tamaños superiores a las 100 pb y menor a 10 kb. Los 30 fragmentos bajo estudio se purificaron mediante un procedimiento de lavado, unión a bolas magnéticas y elución con 30-50 μ L de etanol o agua. Previamente a la utilización del mismo sobre las placas finales, se optimizó el proceso para nueve muestras independientes, buscando la maximización del rendimiento obtenido. Las condiciones finales de mayor rendimiento se obtuvieron manteniendo las especificaciones por defecto del fabricante, pero eluyendo el ADN con etanol a 37 °C, obteniéndose un rendimiento medio del 82%.

Para la preparación de librerías se empleó el kit SureSelect v.4 *Target Enrichment System* (Agilent Technologies, Santa Clara, CA, USA. Tras la creación de librerías, se procedió a la realización de la amplificación por PCR en emulsión (ePCR), y la deposición del ADN unido a bolas magnéticas sobre el *slide* de secuenciación.

Debido al gran tamaño de las librerías generadas para la capacidad de la plataforma de secuenciación, durante la creación de las mismas se tuvieron que dividir los 172 pools de amplificación en 4 pools de secuenciación diferentes, de igual número de casos y controles, para su posterior identificación. Esto se realizó mediante un sistema de 43 códigos de barras que permitiera la distinción de los 43 pools de muestras que componían cada una de las 4 pools de secuenciación, secuenciados en *lines* diferentes de modo que no se confundieran los códigos de barras de cada pool de muestras dentro de cada pool de secuenciación.

3.7. SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS). PLATAFORMAS DE ILLUMINA

3.7.1. Introducción teórica

La tecnología de secuenciación de nueva generación de Solexa (luego adquirida por Illumina), salió al mercado en 2006. Esta tecnología se caracteriza por el empleo de nucleótidos terminadores marcados con moléculas fluorescentes al igual que en la secuenciación de Sanger. Las longitudes de lectura obtenidas son de 150 bases, algo mayores que los secuenciadores SOLID, pero menores a los pirosecuenciadores 454. Sin embargo, la tecnología de Illumina se caracteriza por su capacidad a la hora de realizar secuencias en paralelo (hasta 250 millones de

secuencias). Como resultado es posible obtener hasta 1.3 Gb de datos de secuencia en tiempos similares a los de SOLID.

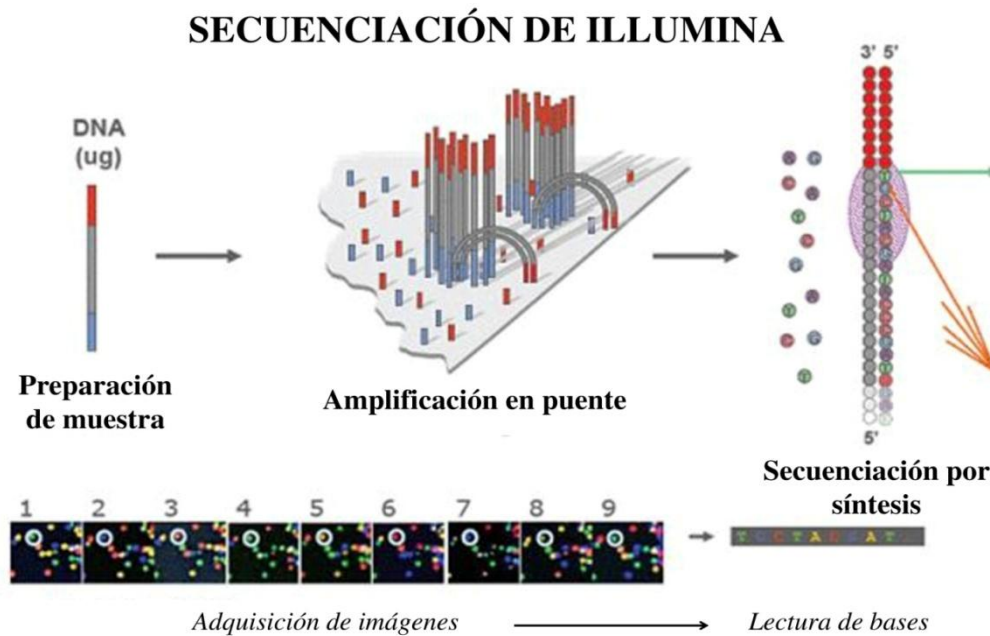
Al igual que para la tecnología SOLID, los primeros pasos consisten en la fragmentación del ADN y ligación de adaptadores para la preparación de librerías.

Las principales diferencias se encuentran en la metodología de amplificación y de secuenciación implicadas. En este caso, la amplificación no tiene lugar en emulsión sino sobre una superficie sólida denominada *flow cell*, donde también tiene lugar la reacción de secuenciación (**figura 3.7**).

- La librería se deposita en la *flow cell* por complementariedad con los adaptadores unidos a ella (P1 y P2) y tiene lugar la **amplificación en puente**. En ella, los adaptadores operan como cebadores y se secuencian las hebras complementarias a la unida a la placa. Tras un posterior lavado, se eliminan todos los fragmentos unidos por enlaces intermoleculares de hidrógeno y se quedan únicamente aquellos sintetizados a partir de los adaptadores de la propia *flow cell*. A partir de entonces, los fragmentos, que contienen ambos extremos, se pliegan en forma de puente para unirse por complementariedad con los adaptadores de la placa, permitiendo la amplificación desde ambos extremos. Este proceso se repite de manera que se forman *clusters* de amplificación (unos 30-50 millones) con aproximadamente 1000 copias en cada uno, en ambos sentidos (desde el adaptador P1 y P2)
- Tras la amplificación tiene lugar la **secuenciación por síntesis**. En este proceso, en primer lugar se cortan los fragmentos que comienzan desde uno de los adaptadores quedando únicamente aquellos unidos a la placa mediante uno de los adaptadores. La reacción de secuenciación comienza desde el otro extremo mediante un cebador de secuenciación. Durante la reacción se ofrecen los cuatro nucleótidos terminadores marcados cada uno con un fluorocromo diferente, de forma similar al método de Sanger. Tras un lavado, se obtienen las imágenes y con ellas la primera base de cada *cluster*. Posteriormente, se elimina el fluorocromo y se desbloquea el Carbono en 3', permitiendo que un nuevo nucleótido pueda extender la cadena de ADN secuenciada, y así, sucesivamente.
- En **secuenciación *paired-end*** (es decir, secuenciando desde ambos extremos de los fragmentos de interés), se repite el proceso anterior pero esta vez rompiendo el enlace desde el otro adaptador de la placa. Para ello, las *flow cell* se despojan de todas las moléculas unidas por complementariedad, tiene lugar la **amplificación en puente** para volver a formar fragmentos unidos desde ambos extremos y, finalmente, se cortan los

fragmentos unidos desde el otro adaptador. El proceso de **secuenciación por síntesis** tiene lugar de nuevo, pero esta vez leyendo los fragmentos desde el otro extremo.

Figura 3. 7. Esquema de la secuenciación de Illumina. Figura modificada de <http://openwetware.org/>



3.7.2. Secuenciación de ARN de muestras de cerebro *post mortem* en esquizofrenia.

Muestras de transcriptoma *post mortem* de giro cingulado de 31 pacientes esquizofrénicos y 26 controles procedentes de la base de datos del Stanley Neuropathology Consortium Integrative Database (<http://www.stanleyresearch.org/brain-research/array-collection/>), apoyado por el Stanley Medical Research Institute (SMRI), fueron secuenciados para su posterior análisis de expresión y variación genética a lo largo de la secuencia de ARN en casos y controles.

Para ello, se llevó a cabo una secuenciación *paired-end* en reads de 75 pb mediante la plataforma Illumina HiSeq, tanto para el conjunto de muestras de individuos enfermos como de sanos, siguiendo las instrucciones del fabricante. De cada muestra, se obtuvieron dos archivos huérfanos de formato FASTQ (por defecto y `_1.fastq` `_2.fastq`), correspondientes a las lecturas en ambos sentidos del proceso *paired-end*, desde los dos adaptadores de secuenciación (**apartado 3.7.1.2**).

El proceso de secuenciación del transcriptoma de las muestras empleadas se realizó en el Columbia Genome Center de la Universidad de Columbia (NY, Estados Unidos), con anterioridad al inicio de la estancia predoctoral que dio lugar a este trabajo.

3.8. ANÁLISIS BIOINFORMÁTICOS

3.8.1. Secuenciación dirigida de ADN de regiones promotoras

3.8.1.1. Mapeo frente al genoma de referencia.

El proceso de mapeo de *reads* procedentes de la secuenciación de promotores mediante la plataforma de SOLiD, se realizó frente a la versión hg19 del genoma humano mediante el programa Bioscope (Applied Biosystems, Carlsbad, CA).

3.8.1.2. Detección y filtrado de variación genética rara en promotores.

Los duplicados fueron eliminados mediante el programa Picard Tools v 1.106, (<http://picard.sourceforge.net>), y las inserciones y deleciones fueron realineadas mediante la función *local realignment* del programa GATK (DePristo y cols., 2011).

Los archivos de alineamiento de salida de Bioscope (en formato SAM) fueron modificados mediante el paquete *Picard Tools v.1.106* (<http://picard.sourceforge.net>) para añadir las informaciones de grupos de lectura, ordenar los archivos, eliminar duplicados e indexar el archivo de salida (en formato BAM, donde la información procedente del archivo SAM se comprime más eficientemente). Estos pasos se realizaron con los programas *AddOrReplaceReadGroups.jar* y *MarkDuplicates.jar*, del paquete Picard Tools.

Los archivos de alineamiento binarios BAM se convirtieron a formato PILEUP mediante el programa SAMtools (Li y cols., 2009). Este formato fue escogido por ser el de entrada para el programa VarScan 2, en el cual se procedió a realizar la detección de SNV, y especialmente recomendado para la detección de variantes en pools (Koboldt y cols., 2012). Éste fue empleado para la detección de variantes con los parámetros seleccionados: profundidad de lectura mínima en una posición genómica (*--min-coverage*), 180; lecturas con mínimas en una posición determinada para la detección de una variante (*--min-reads2*), 4; la calidad mínima de una base en una posición para contar como lectura (*--min-avg-qual*), 0; valor P umbral por defecto para el *variant calling* (*--p-value*), 0,05; umbral de frecuencia mínimo para el alelo de la variante (*--min-var-freq*), 0,01.

La inspección de los resultados reveló la existencia de múltiples posiciones con muchos pools que presentan variantes a frecuencias cercanas a 0,01, considerablemente más bajas que la esperada para los alelos presentes sólo una vez en un pool de los 12 presentes (0,083), 2 por cada uno de los 6 individuos, sugiriendo la presencia de muchos resultados falsos positivos.

Por ello, dos filtros adicionales fueron implementados en el programa R. En primer lugar, se estableció un umbral de frecuencia mínimo de 0,025 para considerar una variante en un pool determinado. En segundo lugar, con el propósito de eliminar posibles falsos positivos por sesgo de secuenciación en un sentido concreto, se estableció un mínimo de 3 lecturas por alelo en cada una de las hebras. Las variantes detectadas por VarScan 2 que no superaron estos filtros fueron eliminadas.

Finalmente, para eliminar variantes en posiciones propensas, por la naturaleza de su secuencia, a contener falsos positivos, se aplicó un filtro posicional comparando la detección de SNV, quitando aquellas posiciones que mostraron las siguientes diferencias en el número de pools en las que presentaban una SNV antes y después de la aplicación de estos dos filtros:

- Si el número total de pools con el alelo raro en el set filtrado fuese de cinco o menos, no se permitió ninguna diferencia. En este caso, las posiciones con alguna diferencia antes y después de la aplicación de los filtros descritos fueron eliminadas.
- En el caso de posiciones con variante en 6-14 pools, se permitió una diferencia en el total de variantes antes y después de los filtros,
- Finalmente, 2 diferencias fueron permitidas si el número de pools mutados en una posición determinada fuese de 15-34.

Igualmente, y dado que el propósito del presente estudio fue la detección de variación genética rara en las regiones seleccionadas, se estableció una frecuencia máxima de 1% para considerar la variante "rara" y, por tanto, incluirla en el conjunto final para analizar.

En primer lugar se estableció un límite en la frecuencia interna (en los 172 pools) que no se podría sobrepasar para considerar esa variante rara. En este sentido, el número máximo de de pools mutados permitido fue de 34. En una distribución binomial, la probabilidad de que aparezca una SNV a una frecuencia de 1% en más de 34 pools es inferior a 0,05%. Por lo tanto, aquellas presentes en más de 34 de los 172 pools, se consideraron comunes y fueron desechadas en ulteriores análisis.

Por otra parte, solo fueron consideradas aquellas variantes cuya frecuencia fuese menor al 1% en la fase 3 del proyecto de los 1000 genomas (1000 Genomes Project Consortium, 2012) y en el proyecto de secuenciación de exomas del NHLBI (ESP6500SI-V2; <http://evs.gs.washington.edu/EVS/>).

3.8.1.3. Anotación funcional de la variación genética detectada

La anotación inicial de las variantes se realizó mediante ANNOVAR (Wang y cols., 2010a). Además, las variantes detectadas fueron clasificadas en función de su potencial regulador en base a las predicciones de RegulomeDB (Boyle y cols., 2012). Tal y como se expone en la **tabla 1.10**, en la introducción, la categoría 1 de funcionalidad está restringida a aquellas con evidencias de actuación como eQTL, por lo que quedarían mayoritariamente excluidas del estudio de variación rara. Por ello, se consideró la categoría 2 como la de máxima evidencia disponible, según esta base de datos, para la funcionalidad en la regulación de la expresión génica. Las variantes fueron clasificadas según su pertenencia a esta categoría, que implica lo siguiente:

- Evidencia experimental mediante secuenciación de cromatina inmunoprecipitada (CHIP-seq) de la ubicación de la variante en una posición que representa algún motivo de unión de proteínas.
- Evidencia experimental de localización de la variante en un lugar accesible del genoma, mediante ensayos de hipersensibilidad a DNasa.
- Ubicación de la variante en algún motivo de unión de factores de transcripción.

De media, un 1,48% de las variantes presentes en el genoma humano están presentes en esta categoría funcional de 2 en RegulomeDB.

Por otra parte, e independientemente de la anotación en RegulomeDB, una segunda clasificación de las variantes anotadas fue realizada de acuerdo a los 15 estados de la cromatina diferentes descritos en proyecto Roadmap (Kundaje y cols., 2015; **tabla 1.9**). Éstos, basados en las modificaciones de 5 de las marcas históricas más importantes en 137 epigenomas humanos de tejidos y líneas celulares diferentes, fueron descargados de la base de datos del navegador del UCSC para tres tejidos de importancia en el desarrollo de la esquizofrenia: Cerebro fetal (masculino y femenino), giro angular (*angular gyrus*) y corteza prefrontal medial (*mid frontal lobe*).

- Los datos de cerebro fetal, relativos a hombre y mujer, fueron elegidos teniendo en cuenta la implicación de los genes cuyos promotores se secuenciaron en el presente trabajo en diferentes trastornos del neurodesarrollo.
- Los dos tejidos adultos, giro angular y corteza prefrontal medial, se eligieron en base a los resultados del último GWAS en esquizofrenia, y de mayor tamaño, realizado hasta la fecha (Schizophrenia Working Group of the Psychiatric Genomics Consortium,

2014), en el que se destacaba la importancia de estos dos tejidos en la regulación de la expresión génica a partir del análisis de la modificación histónica H3K27ac.

3.8.2. Secuenciación de transcriptoma de giro cingulado en esquizofrenia

3.8.2.1. Mapeo frente al genoma de referencia.

Para el proceso de mapeo de *reads* procedentes de la secuenciación de la plataforma de Illumina, se utilizó el programa STAR aligner, dada la mejora en precisión y rapidez de mapeo en comparación con las últimas versiones de su homólogo para ARN más utilizado, Tophat2 (Dobin y cols., 2013). Para ello, las lecturas procedentes del *Genome Analyzer* de la plataforma de Illumina fueron procesadas y alineadas al genoma de referencia mediante *STAR Aligner* v2.3 utilizando un procedimiento explicado a continuación.

Los archivos de secuencia del genoma humano (versión hg19) para cada cromosoma, en formato *fasta*, fueron descargados del repositorio de la base de datos del navegador del UCSC (<http://hgdownload.soe.ucsc.edu/downloads.html>), fueron unidos y ordenados de forma compatible con las especificaciones del programa STAR. El alineamiento fue llevado a cabo utilizando los parámetros por defecto del protocolo y de *STAR-2 pass*, recomendando por el *Broad Institute* para el estudio de datos de RNA-seq (<https://www.broadinstitute.org/gatk/guide/article?id=3891>).

3.8.2.2. Estudio de la expresión diferencial en giro cingulado en esquizofrenia

Los archivos de alineamiento de salida del STAR aligner (formato SAM, del inglés *Sequence Alignment Map*) fueron inicialmente ordenados mediante el programa samtools v.1.2 (Li y cols., 2009) y se utilizaron como archivo de entrada en *htseq-count*, una herramienta desarrollada en el paquete HTSeq v 0.6.1 (Anders y cols., 2014) que procesa los datos de RNA-Seq para el análisis de la expresión diferencial contando el solapamiento de lecturas con los genes cuyas coordenadas se incluyen en un archivo de anotación. Los vectores de conteo de lecturas se obtuvieron para cada muestra y, posteriormente, se fusionaron en una matriz final con la información de todas las lecturas por gen y por muestra. Como archivo de anotación se utilizó el de la base de datos de Gencode v19 (*Gencode v19 comprehensive annotation*) obtenido a partir del navegador del UCSC. Éste fue utilizado para anotar los genes y sus distintas isoformas. Dada la superposición de muchas isoformas y la dificultad de distinguir a cuales de las mismas se atribuyen las lecturas mapeadas, en caso de solapamiento de

coordenadas genómicas, los genes se fusionaron en *loci* únicos mediante la herramienta *cuffmerge* del software de análisis de ARN *Cufflinks* (Trapnell y cols., 2012). De este modo, el archivo final de anotación que sirvió para la creación del vector de conteo de lecturas mediante *htseq-count* consistió en 48099 *loci* no solapantes que cubren en su totalidad 1559 Mb del genoma humano, obtenidos del total de transcritos de Gencode v 19 (<http://www.gencodegenes.org/stats/archive.html#a19>).

Posteriormente, a partir de la matriz de conteo de lecturas se creó otra en valores de RPKM (lecturas por kilobase de longitud y millón de lecturas mapeadas), de manera que se pudieran comparar los valores de expresión entre las distintas muestras (independientemente de las diferentes coberturas en ellas) y distintos genes (independientemente de la diferente longitud entre ellos) mediante la siguiente transformación:

$$RPKM = \frac{N_{lecturas}}{L_{exónica}} \times \frac{1000000}{N_{lecturas\ en\ muestra}}$$

En esta expresión $N_{lecturas}$ representa el total de lecturas en un gen determinado, $N_{lecturas\ en\ muestra}$ el total para una muestra determinada y $L_{exónica}$ la longitud de la secuencia exónica del gen. Los valores de RPKM se calculan para cada uno de los 48099 *loci* y cada una de las 57 muestras.

Los *loci* con valores de RPKM < 0,3 en todo el conjunto de datos de expresión se eliminan, ya que este umbral se ha definido previamente como un punto de corte fiable para distinguir la expresión de ruido de fondo (Ramsköld y cols., 2009). Este filtro redujo el número de *loci* iniciales de 48099 a 22386 *loci* con niveles de expresión detectables en las muestras analizadas.

Los valores expresión, en conteo de lecturas por gen, de los 22386 *loci* con expresión detectable se exportaron a un nuevo archivo que sirvió de *input* para los *software* EdgeR v.3.13.0 (Robinson y cols., 2010) y DESeq v.1.22.0 (Anders y cols., 2010), programas basados en lecturas por gen para analizar las diferencias de expresión génica entre casos y controles. Ambos *software* son implementados en el paquete estadístico R para determinar los genes expresados diferencialmente (DE). Solo los resultados solapantes en DESeq y EdgeR fueron considerados, generando una lista de 1876 genes DE en esquizofrenia respecto a los valores en controles.

Con el fin de explorar el significado biológico e interpretar las diferencias en la expresión génica, se realizó un análisis de enriquecimiento en grupos de genes concretos mediante la herramienta GSEA, del inglés *Gene Set Enrichment Analysis* (Subramanian y cols., 2005). Este análisis fue realizado para comprobar el enriquecimiento de las diferencias de

expresión génica en grupos de genes de estudios previamente publicados y referidos a conjuntos de genes coexpresados a lo largo del desarrollo (Hawrylycz y cols., 2012) y en el transcriptoma prenatal (Miller y cols., 2014). Igualmente, también se estudió el enriquecimiento en conjuntos de genes específicos de tipos celulares concretos de cerebro, también procedentes de estudios previos mediante microarrays de expresión (Cahoy y cols., 2008) o RNA-seq (Zhang y cols., 2014a). Finalmente, sets de genes relacionados con la sinapsis y descritos en la base de datos de SynaptomeDB (Pirooznia y cols., 2012) también se utilizaron para estudiar el enriquecimiento en los datos de expresión. Para cada prueba, como archivos de entrada se utilizaron la matriz de expresión en RPKM (22386 *loci*), un archivo de información fenotípica de las muestras y otro con los genes incluidos en cada uno de los sets utilizados. Para el análisis de enriquecimiento se requirieron un mínimo de 10 genes en el set presentes en la matriz de expresión para considerar dicho set, se realizaron 1000 permutaciones de fenotipo (caso o control) y se utilizaron los parámetros *weighted statistic enrichmenty signal2noise* para el procedimiento estadístico.

La herramienta GSEA también se utilizó para evaluar el enriquecimiento en los sets indicados previamente utilizando la matriz de expresión solo con genes no correlacionados con los niveles de medicación en casos y utilizando la matriz de expresión únicamente con genes altamente correlacionados con niveles de medicación, tal y como se explica posteriormente.

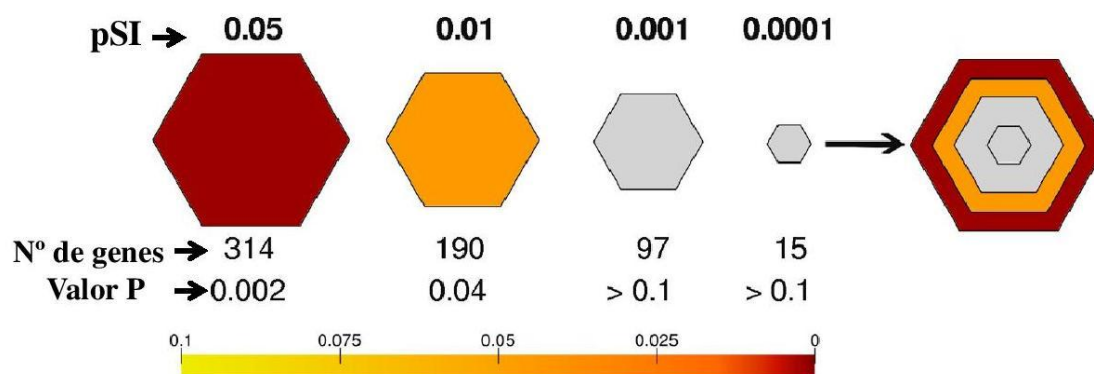
Además, también se analizó el enriquecimiento de los 1876 genes DE en aquellos genes específicos de tipos celulares cerebrales y regiones cerebrales durante el neurodesarrollo mediante la herramienta de análisis de expresión específica de tipo celular CSEA, del inglés *Cell-type Specific Enrichment Analysis* (Dougherty y cols., 2010, Xu y cols., 2014), de libre acceso y disponible en la red (<http://genética.wustl.edu/jdlab/csea-tool-2/>). Mediante esta herramienta, se analiza el enriquecimiento en la expresión espaciotemporal de los genes diferentemente expresados en el presente trabajo mediante comparación con datos de expresión de la base de datos de BrainSpan (<http://www.brainspan.org/>) para distintas líneas celulares y regiones cerebrales en distintas etapas del neurodesarrollo. El resultado del análisis se representa en hexágonos concéntricos para cada línea celular o región cerebral analizada, donde el tamaño de los mismos es proporcional al número de genes en cada índice de probabilidad de especificidad (pSI) y el color del relleno representa el valor de significación del enriquecimiento (**figura 3.8**)

El análisis de CSEA también se realizó en genes DE sobreexpresados (1133) y subexpresados (743), por separado.

La herramienta CSEA también se utilizó para evaluar el enriquecimiento en las líneas celulares y regiones cerebrales durante el neurodesarrollo utilizando solamente el conjunto de genes DE no correlacionados con los niveles de medicación en casos o bien utilizando el

conjunto de genes DE únicamente con aquellos altamente correlacionados con niveles de medicación, tal y como se explica posteriormente.

Figura 3. 8. Esquema de la representación de los resultados de CSEA. Figura adaptada de Xu y cols. (2014).



3.8.2.3. Estudio de la correlación entre expresión génica y medicación administrada a lo largo de la vida en los individuos esquizofrénicos.

La influencia de la exposición a la medicación administrada a los 31 esquizofrénicos durante su vida en los niveles de expresión génica podría ser un factor de confusión a la hora de evaluar las diferencias de expresión entre casos y controles, en el contexto de una consecuencia natural de la enfermedad. Los valores normalizados de medicación administrada durante la vida de los individuos afectos se integraban como parte de la información de las muestras proporcionadas por la Fundación Stanley.

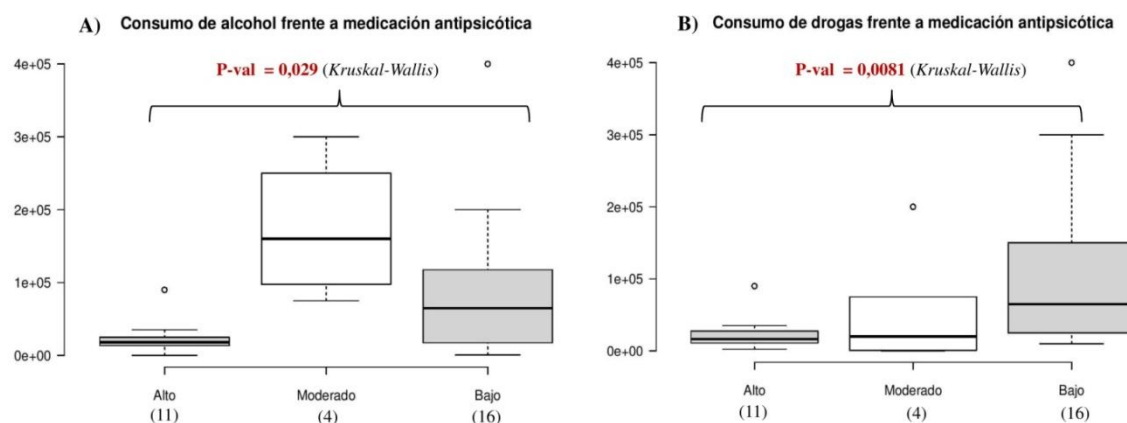
En este sentido, se calcularon las correlaciones de Pearson entre los valores de expresión génica y los de exposición a medicación en las 31 muestras de casos (los controles no estaban expuestos a medicación), para cada uno de los 22386 *loci* con expresión detectable (RPKM > 0,3).

La totalidad de los genes expresados (22386 *loci*), así como los genes DE (1876 *loci*), se dividieron en tres grupos diferentes para sus posteriores análisis. Se consideró que un valor de significación de la correlación de Pearson corregido FDR-p > 0,05 era indicativo de un locus no correlacionado con la medicación. Igualmente, para profundizar en la influencia de la medicación en la expresión génica, también se consideró el subconjunto de *loci* con correlación de Pearson corregida FDR-p < 5 x 10⁻⁵ como el grupo de genes fuertemente correlacionados con la medicación. De esta forma, para comparaciones durante análisis de expresión y variantes genéticas, se consideraron tres listas de genes a partir de aquellos con expresión detectable:

todos los genes con expresión detectable (todos los 22386 *loci*), genes fuertemente correlacionados con la medicación (18854 *loci* con $\text{FDR-p} < 5 \times 10^{-5}$) y genes no correlacionados con la medicación (2104 *loci* con $\text{FDR-p} > 0,05$). Del mismo modo, para el análisis de genes DE, también se consideraron los mismos grupos de genes: todos los genes DE con expresión detectable (1876 *loci*), genes DE fuertemente correlacionados con la medicación (1585 *loci*) y genes DE no correlacionados con la medicación (174 *loci*).

Además de la medicación, otras variables como el alcohol y el consumo de drogas podrían, igualmente, influir en las diferencias de expresión y dar lugar a falsos positivos en los datos de diferencias de expresión. Sin embargo, se halló una asociación significativa de los niveles de consumo de drogas (diferenciados entre alto, medio y bajo o social) y los de consumo de alcohol (de igual modo divididos en niveles alto, medio y bajo o social) con las dosis de medicación administrada a lo largo de la vida. De esta forma, mediante el análisis de correlación de la medicación con la expresión génica, se tuvieron en cuenta (por asociación indirecta) las posibles diferencias en el consumo de alcohol y drogas de abuso entre casos y controles estudiados (**figura 3.9**).

Figura 3. 9. Asociación entre medicación administrada de los pacientes esquizofrénicos (31) cuyas muestras se analizaron y el consumo de alcohol (A) y drogas de abuso (B). Los niveles de consumo de ambas se dividieron en tres grupos (Alto, moderado y bajo) para los descritos por el SMRI. Para el estudio de la relación entre los niveles de medicación y ambos consumos, se empleó un test no paramétrico de Kruskal-Wallis, comparándose los valores de expresión en RPKM (en el eje de ordenadas).



3.8.2.4. Detección y filtrado de variación genética en transcriptoma de cerebro en esquizofrenia.

Los archivos de alineamiento de salida del STAR aligner (en formato SAM) fueron modificados mediante el paquete *Picard Tools v.1.109* (<http://picard.sourceforge.net>) para añadir las informaciones de grupos de lectura, ordenar los archivos, eliminar duplicados e

indexar el archivo de salida en formato BAM. Estos pasos se realizaron con los programas *AddOrReplaceReadGroups.jar* y *MarkDuplicates.jar*, del paquete Picard Tools.

Los archivos de salida en formato BAM, para las muestras analizadas, se analizaron en GATK (DePristo y cols., 2011) siguiendo el protocolo recomendado por el *Broad Institute* para el análisis de variantes a partir de datos de RNA-seq (<https://www.broadinstitute.org/gatk/guide/article?id=3891>). Brevemente, en primer lugar se aplicó la herramienta *Split N Cigar Reads*, desarrollada específicamente para el análisis de RNA-seq y cuyo objetivo es el corregir los errores producto del alineamiento en lugares de *splicing* y de unión entre exones. A continuación, los archivos se sometieron a realineamiento, recalibración de las bases y detección de variantes (*variant calling*) mediante las herramientas de GATK *Indel Realigner*, *Base Recalibration*, y *Haplotype Caller*, respectivamente. Las bases de datos de variantes puntuales (*dbSNP138*) y de indels (*Mills and 1000G gold standard indels*) se utilizaron como "lugares conocidos" para recalibrar los valores de calidad en todos los archivos de BAM. La herramienta *Haplotype Caller* se utilizó en lugar de *Unified Genotyper*, por su mayor precisión en la detección de inserciones y deleciones, y los argumentos *recoverDanglingHeads* y *dontUseSoftClippedBases* fueron empleadas a fin de reducir al mínimo los falsos positivos y falsos negativos. El valor mínimo de calidad *Phred* se redujo para el *variant calling* desde el valor predeterminado de 30 hasta 20, siguiendo los criterios del *Broad Institute* para datos de RNA-seq. Para cada muestra se obtuvo un archivo en formato VCF con la información de la variante tras el procedimiento de *variant calling* y, posteriormente, se filtraron las variantes detectadas con sesgo en la detección en alguna de las hebras (valores *Fisher Strand* (FS) > 30) y aquellas de baja confianza (Quality by Depth values (QD) < 2,0). Además, también se eliminaron las variantes pertenecientes a grupos de SNP (*SNP clusters*) en intervalos de menos de 35 pb entre ellas.

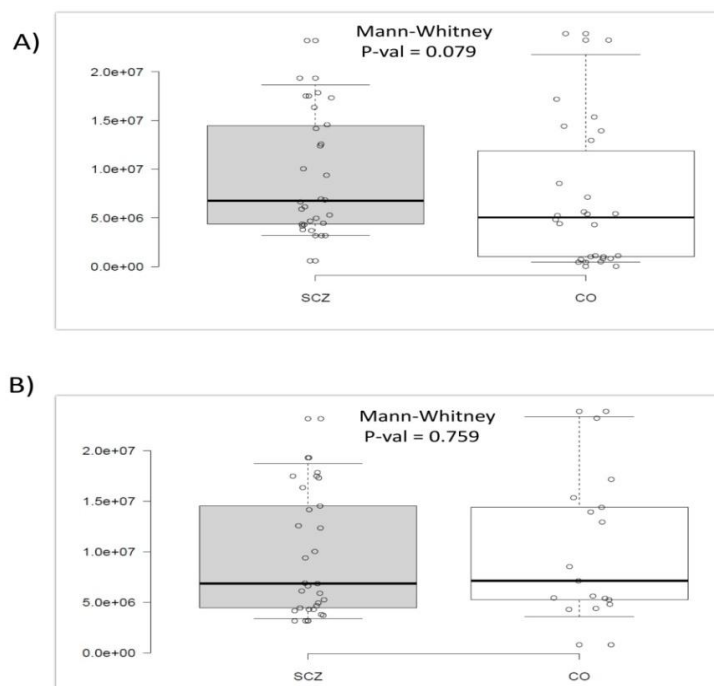
3.8.2.5. Detección de variación genética en genes expresados consistentemente en giro cingulado

Con el propósito de mejorar la calidad de detección de las variantes y restringirlas a genes expresados de forma ubicua, se utilizó un umbral posicional de profundidad de lectura (DP > 10) en todas las muestras para excluir posiciones con un valor de DP < 10 en alguno de los individuos, y de ese modo, sólo seleccionar aquellas variantes en posiciones genómicas que se expresan de forma consistente en todas las muestras.

Para ello, se empleó el programa *Genomecov* del paquete de análisis Bedtools (Quinlan y cols., 2014), para el cálculo de la cobertura en los archivos BAM procesados, sin duplicados y

ya recalibrados. Como archivo de anotación se utilizó el de Gencode v19 (*Gencode v19 comprehensive*) fusionado con *cuffmerge*, tal y como en el caso del análisis de expresión. Por último, se utilizó el programa *IntersectBED* de Bedtools para extraer las posiciones que satisfacen el criterio de DP. A fin de estudiar las variantes genéticas detectadas en relación con los perfiles de expresión estudiados, la variación genética se restringió a la presente en los genes con expresión detectable previamente descritos (22386 *loci* con RPKM > 0,3). En total, 1,35 Mb correspondientes a posiciones en genes altamente expresados, y de forma ubicua en el conjunto de muestras, fueron objeto de estudio de su variación genética.

Figura 3. 10. Diagrama de cajas y bigotes para la comparación de la cobertura del transcriptoma (a DP > 10) de A) la totalidad de las 31 muestras de casos y 26 de controles o B) las 29 muestras de casos y 17 de controles empleadas en el análisis de variantes. En cada caso, se realizaron test de Mann-Whitney para el estudio de coberturas en los grupos de casos y controles, comparándose los valores de DP en cada caso (en el eje de ordenadas).



El estudio de variantes genéticas se limitó a 29 muestras de ARN de esquizofrenia y 17 de controles de los 31 y 26 totales, respectivamente. 2 casos y 9 controles se excluyeron del análisis debido a su menor cobertura en comparación con el resto de las muestras (**figura 3.10; ANEXO I**), lo que a su vez reduciría ostensiblemente el número de posiciones genómicas analizadas con DP > 10 en todas las muestras de haberlas incluido. Todos los resultados del análisis de expresión fueron repetidos incluyendo solo las muestras utilizadas en el análisis de variantes, con el propósito de descartar la posibilidad de que los resultados encontrados en el

análisis de expresión se debieran en algún modo a la presencia de las muestras eliminadas de baja cobertura.

3.8.2.6. Anotación funcional de variantes en transcriptoma de giro cingulado

La anotación de las variantes se realizó mediante ANNOVAR (Wang y cols., 2010a). Los archivos VCF filtrados son convertidos a formatos compatibles con esta herramienta mediante el programa *convert2annovar.pl*, fusionados en un único archivo y anotado mediante *table_annovar.pl*.

Con el objetivo de estudiar la variación rara presente en genes altamente expresados de las muestras analizadas, se eliminó la variación genética previamente descrita con frecuencia superior a 0,01 en cualquiera de las poblaciones de la fase 3 del proyecto de los 1000 genomas, del *exome sequencing project* (ESP).

3.9. ANÁLISIS ESTADÍSTICOS

3.9.1. Estudio de la variación genética en UC435 de TCF4

Para el cálculo de potencia estadística en la detección de variantes genéticas de susceptibilidad a lo largo del elemento UC435, se empleó el programa GPower 3 (Faul, 2007).

3.9.2. Estudio de la variación genética en HAEs de NPAS3

Haploview 4.2 (Barrett y cols., 2005) se utilizó para analizar el DL presente entre las variantes seleccionadas, para chequear el equilibrio de Hardy-Weinberg en cada variante genotipada y para realizar las pruebas de asociación a nivel alélico individual para cada variante.

El efecto acumulado de las diferentes variantes en el riesgo al desarrollo de esquizofrenia se estudió mediante test de asociación implementados en SKAT (Ionita-Laza y cols., 2013; Lee y cols., 2012a; Wu y cols., 2011) un programa diseñado para el entorno R. Para la preparación de los archivos con la información del genotipo de las variantes analizadas, se reformatearon utilizando PLINK (Purcell y cols., 2007). SKAT también se utilizó para la realización de cálculos de potencia, suponiendo en ellos una prevalencia de la enfermedad de 0,01, y una longitud de secuencia igual a la suma de la longitud HAEs estudiados (6489 pb).

Dado que la arquitectura genética real del riesgo potencial ejercido por estas variantes en el contexto de la esquizofrenia es desconocida, se utilizaron diferentes pruebas para obtener la máxima potencia bajo diferentes escenarios. En cuanto a la dirección del efecto, tres pruebas estadísticas diferentes fueron empleadas.

- Pruebas de carga en las que se supone que todas las variantes funcionales ejercen su efecto en la misma dirección, aumentando el riesgo.
- Alternativamente, las pruebas de análisis del componente de varianza se emplearon para tener en cuenta la dirección del efecto ejercido por cada variante. Bajo este hipotético escenario de coexistencia de variantes de riesgo y de protección, la utilización de este test mejoraría la potencia estadística del análisis.
- Un último test considerado fue el *optimal unified test*, diseñado para la combinación del test de carga y de varianza (Lee y cols., 2012a).

A su vez, respecto a la frecuencia de las variantes de susceptibilidad analizadas, se utilizaron tres tipos de test (Ionita-Laza y cols., 2013), considerando los efectos ejercidos por:

- Únicamente las variantes genéticas raras.
- Todas las variantes pero sobrerrepresentando la importancia de las raras frente a las comunes.
- Todas las variantes asumiendo efectos combinados de variantes raras y comunes.

Los tres tipos distintos de pruebas estadísticas considerando la frecuencia de las variantes fueron corridos mediante las funciones *SKAT* o *SKAT_CommonRare*, con los argumentos *r.corr* y *test.type* alterados en función de la prueba. Los valores por defecto para el resto de parámetros no se modificaron, así como el umbral considerado por el propio programa para la asignación de "variante rara" definido como $1/\sqrt{(2N)}$, donde *N* representa el tamaño muestral empleado.

Además de los análisis de variantes individuales y conjuntamente, se realizó un análisis de asociación por comparación de la frecuencia de haplotipos entre casos y controles mediante una prueba de ratio de probabilidad (*likelihood ratio tests*). Los haplotipos existentes en el conjunto de variantes analizados fueron inferidos mediante el algoritmo de esperanza de maximización (*Expectation-Maximization*) del programa Unphased 3.1.7 (Dudbridge, 2008), con un procedimiento de "ventanas deslizantes" (del inglés *sliding windows approach*) de tamaño de ventana de 2 - 6. Este análisis se restringió a las variantes de frecuencia superior al 1% en la base de datos de los 1000 genomas. Los haplotipos raros, definidos como aquellos de frecuencia inferior a 1%, se agruparon antes de la prueba. La significación de los resultados se calculó tras corrección por 1000 permutaciones de los valores fenotípicos.

3.9.3. Estudio de la variación genética en regiones promotoras

La detección de falsos positivos se estimó por comparación de los resultados de resecuenciación y genotipado, suponiendo la no existencia de errores en la determinación del genotipo. Los pocos SNV que no superaron esta validación se excluyeron para los análisis relativos al estudio de la variación genética en las secuencias promotoras analizadas. Dado que la tasa de falsos positivos fue muy baja, el resto de SNV de categorías funcionales diferentes a 2 (y por ello, menos propensas a presentar falsos positivos) no probados fueron considerados como SNV verdaderos.

La proporción de SNV raras en las diferentes categorías funcionales consideradas se comparó entre casos y controles mediante prueba exacta de Fisher de dos colas o la prueba chi-cuadrado (χ^2) de Pearson, empleando la corrección de Yates cuando la frecuencia esperada de alguno de los grupos era inferior a 0,05. En los casos en donde las tablas de comparación de proporciones 2x2 contuvieran algún 0, se empleó el método de regresión logística de estimación de máxima verosimilitud penalizada de Firth (Wang, 2014) para el cálculo de significación estadística, OR e intervalos de confianza mediante el paquete de R *logistf*.

3.9.4. Estudio del transcriptoma de giro cingulado en esquizofrenia

3.9.4.1. Análisis de datos de expresión mediante GSEA y CSEA

Los valores p de significación para cada ensayo realizado de enriquecimiento en GSEA fueron corregidos por múltiples comparaciones mediante el procedimiento de *False discovery rate* (FDR) de Benjamini- Hochberg (Benjamini y cols., 1995).

Para evaluar la asociación estadística de los análisis implementados en CSEA se realizan test hipergeométricos, y estos se corrigen por comparaciones múltiples mediante el procedimiento de FDR de Benjamini- Hochberg (Benjamini y cols., 1995), como parte de los resultados provistos por la herramienta.

3.9.4.2. Análisis de correlación entre medicación y expresión génica

Los valores P obtenidos para cada correlación entre los niveles de medicación en casos y la expresión génica de cada uno de los 22386 *loci*, basándose en un muestreo aleatorio con reemplazo, se obtuvieron utilizando el procedimiento de remuestreo por *bootstrap* (Efron y Tibshirani, 1993). De esta manera, utilizando la calibración del *bootstrap* también se construyeron intervalos de confianza, permitiendo estimar la distribución de las correlaciones de muestreo después de 500 ensayos en el remuestreo, y se calcularon los valores p basándose en

los intervalos de confianza construidos para cada caso. Estos análisis fueron implementados en el entorno R.

Para la corrección por múltiples test, se utilizó el procedimiento de FDR de Benjamini-Yekutieli (Benjamini y cols., 2001), ya que esta corrección mejora las estimaciones del FDR de Benjamini-Hochberg (Benjamini y cols., 1995) bajo formas de dependencia. En el presente estudio, el cálculo de correlación entre los diferentes vectores de expresión génica y el mismo vector de medicación a lo largo de la vida (dado que es el mismo vector independientemente de los genes) dibuja un escenario en el que la corrección por FDR bajo condiciones de dependencia es más apropiada. Estos valores se calcularon en R mediante la función *p.adjust* y el argumento "BY".

3.9.4.3. Variación genética en módulos de coexpresión

Para probar el enriquecimiento de los genes mutados en cada uno de los módulos de coexpresión destacados se utilizó la prueba chi-cuadrado (χ^2) de Pearson, empleando la corrección de Yates cuando la frecuencia esperada de alguno de los grupos era inferior a 0,05.

En el caso de la prueba de enriquecimiento de variantes en casos en comparación con controles en alguno de los módulos destacados se empleó el test exacto de Fisher de dos colas. Todos los valores P se ajustaron por múltiples comparaciones mediante la corrección por FDR de Benjamini y Hochberg (Benjamini y cols., 1995), mediante la función *p.adjust* (con el argumento BH) en el entorno R. Para el análisis de variantes genéticas en genes de módulos sobreexpresados en esquizofrenia se realizó un test de tendencia de Cochran-Armitage de las variantes por individuo en casos y controles.

Las pruebas estadísticas χ^2 y de Cochran-Armitage fueron realizadas mediante el software WINPEPI (Abrahamson, 2011).

3.9.4.4. Estudio de la influencia de la variación genética rara de genes altamente expresados en los niveles de expresión génica.

La importancia del diseño experimental como herramienta estadística en el contexto de estudios de expresión ha sido destacada anteriormente (Kerr y Churchill, 2001). Con el fin de probar la contribución de las variantes genéticas a las diferencias observables en los valores de expresión génica, se realizó un diseño de experimentos para descomponer la varianza observada en los valores de expresión génica (Y) en la matriz de expresión en la contribución de tres

variables diferenciadas (gen, fenotipo y presencia/ausencia de variante en el gen). Este análisis se realizó para tres escenarios diferenciados: En todos los genes mutados con expresión detectable (623 genes), únicamente los genes mutados pertenecientes a módulos de coexpresión sobreexpresados en esquizofrenia (11 genes) y en aquellos genes mutados a pertenecientes a módulos de coexpresión subexpresados en esquizofrenia (345).

El correspondiente análisis de la varianza (ANOVA de tres vías o factores) del diseño de experimentos asociado se realizó de acuerdo al siguiente modelo de regresión matemático:

$$Y = \mu + \alpha_{p=1,2} + \beta_{v=1,2} + \gamma_{g=1,2,\dots,77} + \alpha\beta + \alpha\gamma + \beta\gamma + \alpha\beta\gamma + \varepsilon$$

Aquí, Y es el valor de expresión de cada celda de la matriz de expresión en el conjunto de datos (correspondiente a la matriz de expresión de un gen en una muestra específica); μ es el valor de la expresión basal; α es la variable "fenotipo", que toma valores de $p = 1, 2$ en los casos y controles, respectivamente; β es la variable "variante" que toma valores de $v = 1, 2$ en ausencia o presencia de la variante, respectivamente; γ es la variable "gen" que toma valores de g de 1 a N (donde N es el número de genes mutados con expresión detectable (623) o el número de genes de los módulos de coexpresión sobreexpresados (11) o subexpresados (345) en esquizofrenia).

La técnica estadística de ANOVA para el análisis de la varianza se ha empleado previamente en el análisis de datos de expresión (Fernandes y cols., 2013; Oleksiak y cols., 2002). Esta técnica calcula la significación utilizando el cociente F, o F de Fisher-Snedecor. Este se genera, para cada uno de los tres factores considerados, dividiendo el efecto debido a la pertenencia de los grupos (variante entre grupos) respecto a la dispersión debida al azar (varianza dentro de los grupos). Si sobrepasa cierto valor crítico, entonces se puede afirmar que el efecto observado es demasiado grande como para poder ser explicado por azar y que, por tanto, no todos los grupos estudiados tienen la misma media. De este modo, si el estadístico F es significativo en alguno de los tres factores del modelo, significará que éste contribuye de forma significativa a los valores de expresión de la matriz de expresión. Por lo tanto, para valorar la influencia de la variación genética en los datos de expresión, se analizó la contribución de la presencia o ausencia de variantes genéticas a la varianza existente en la matriz de expresión, en conjunción con la contribución de las otras variables analizadas (fenotipo y genes).

Finalmente, el diseño de experimentos se repitió en los mismos tres escenarios descritos, pero con los valores de expresión normalizados por gen, de forma que se eliminan las

diferencias de expresión entre genes. El propósito de ello es dilucidar si una hipotética contribución de la presencia/ausencia de variantes genéticas en la varianza de la matriz de expresión no normalizada es significativa debido a las diferencias entre genes (esto es, que las variantes tienden a estar dentro de genes mayor o menormente expresados), o debido a la implicación directa de la variante en la alteración de los niveles de expresión. Esta normalización se realizó dividiendo cada valor de expresión en cada gen por la suma de la expresión génica en todas las muestras en ese gen, y multiplicando posteriormente por un valor constante k ($k = 10,000$) para así generar una matriz de expresión normalizada preservando las diferencias de expresión dentro de las muestras de un mismo gen, pero eliminando la variabilidad en la expresión génica entre los diferentes genes analizados. Todos los análisis de ANOVA fueron implementados en el entorno R, utilizando el paquete estadístico *nlme*.





4. RESULTADOS



ESTUDIO I. VARIACIÓN GENÉTICA RARA EN ELEMENTOS CONSERVADOS DE *TCF4* DE POTENCIAL REGULADOR

I. 1. Introducción

TCF4 es un factor de transcripción implicado en el desarrollo del sistema nervioso central y fuertemente asociado con la esquizofrenia, algo demostrado en los estudios de GWAS (Steinberg y cols., 2011a). La haploinsuficiencia del gen *TCF4* es la responsable de la aparición del síndrome de Pitt-Hopkins, una enfermedad caracterizada, entre otras cosas, por fenotipos severos como discapacidad intelectual, alteraciones anatómicas en la boca e hiperventilación. Esta haploinsuficiencia se ha visto causada por mutaciones *nonsense*, que originan codones de parada prematura, mutaciones *missense*, translocaciones o deleciones (Whalen y cols., 2012).

Bejerano y cols. (2004) describieron la existencia de 481 elementos ultraconservados (UCEs) a lo largo del genoma humano, que se definen como secuencias de al menos 200 pb con 100% de homología en el genoma de humano, ratón y rata. Subsiguientes validaciones experimentales de muchos de estos UCEs mediante ensayos de transgénesis en embriones de ratón revelaron que la mayoría de los UCEs no codificantes son elementos reguladores implicados en la expresión génica durante el desarrollo. Uno de los elementos analizados con función de *enhancer* de la expresión génica fue UC435, un UCE enteramente no codificante y cuya secuencia de 227 pb de longitud se sitúa en una región intrónica de *TCF4*. UC435 potencia la expresión en el rombencéfalo, mesencéfalo y tubo neural (*VISTA enhancer browser* (ID: hs376); <http://enhancer.lbl.gov>). De hecho, mediante una inspección visual desde el navegador de UCSC, el elemento es también diana de la proteína p300, una acetiltransferasa de histonas vinculada a la potenciación de la expresión en la línea celular de neuroblastoma (ENCODE Project Consortium, 2012), lo que proporciona una evidencia adicional de la relevancia de su papel como regulador de la expresión génica.

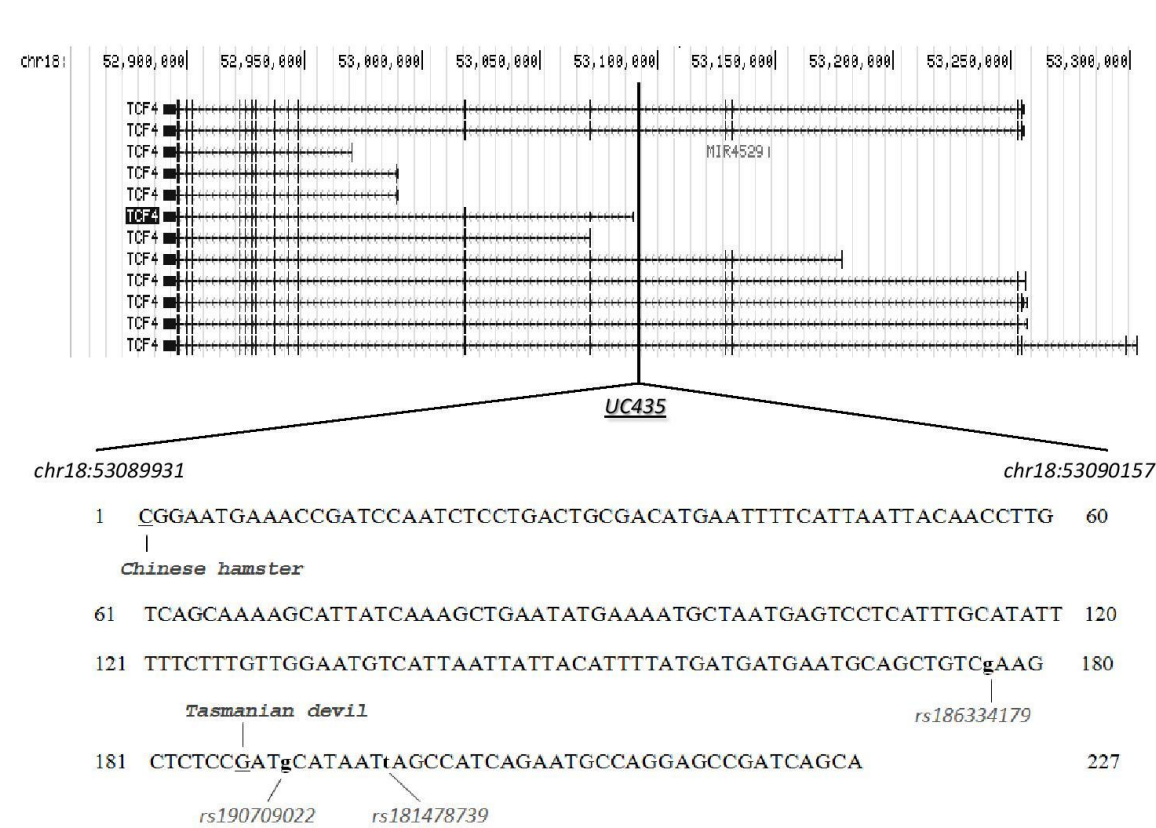
TCF4 presenta, además, otro UCE de los 481 descritos por Bejerano y cols. (2004): UC436. Sin embargo, a diferencia de UC435, casi la mitad de UC436 implica un exón codificante de *TCF4*. Asimismo, no hay evidencia experimental del posible papel de UC436 como potenciador o *enhancer* de la expresión génica. Debido a esto, el presente estudio se centró en la búsqueda de alelos muy penetrantes a lo largo del elemento UC435 de *TCF4*, en un contexto de posible implicación de esas supuestas variantes en la alteración de la normal expresión génica de *TCF4* y su efecto en la aparición de la esquizofrenia.

I. 2. Estudio de la secuencia de UC435 en mamíferos y de la variación genética descrita en población humana.

Se propone el estudio del elemento UC435 por ser completamente intrónico y, por tanto, de mayor interés en lo que respecta a sus potenciales propiedades en la regulación de la expresión génica de *TCF4*.

Tras un análisis pormenorizado de la secuencia de UC435 en otros mamíferos, se reveló que tan sólo existe un único cambio en las secuencias de referencia de 22 mamíferos placentarios, correspondiente a la posición 1 de UC435 en hámster chino, lo que sugiere que los bordes del elemento son aproximados, pero en cualquier caso refuerza la ultraconservación de este elemento, haciéndola extensible a 22 mamíferos adicionales. Por extensión de la comparación de los genomas de tres marsupiales, se comprobó que dos de ellos presentan idéntica secuencia de UC435 y uno, que se corresponde al diablo de Tasmania, presenta un solo cambio (**figura 4.1**).

Figura 4. 1. Variación genética en UC435. Con el nombre de la especie se representan las 2 variantes encontradas en 22 mamíferos placentarios y 3 marsupiales analizados. Las variantes señaladas con rs pertenecen a dos individuos de Kenya (rs190709022), una mujer japonesa (rs186334179) y un hombre también japonés (rs181478739), tras la búsqueda en la base de datos de los 1000 genomas.



Curiosamente, la búsqueda en los datos del Proyecto 1000 Genomas (1000 Genomes Project Consortium, 2012) reveló la existencia de 3 variantes poco comunes en UC435, dos detectadas en una sola ocasión (rs186334179, en una mujer japonesa de Tokio, y rs181478739, también en un japonés de Tokio) y la otra (rs190709022), presente en dos sujetos de la etnia *Luhya*, en Kenia (**Figura 4.1**).

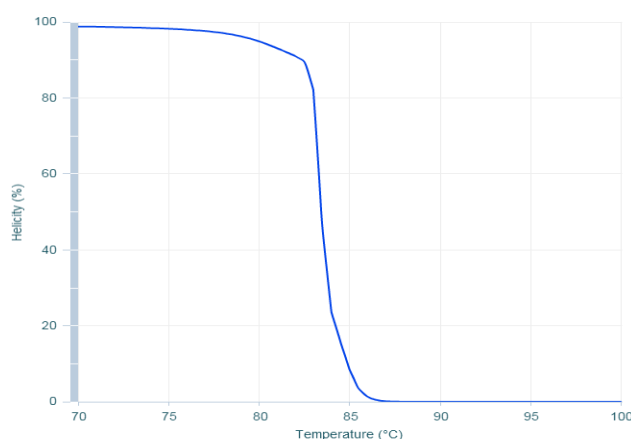
I. 3. Análisis de la variación genética en UC435 en esquizofrenia mediante *High Resolution Melting Analysis* (HRMA).

El fragmento a amplificar presentó una longitud de 284 pb en el UC435 (chr18:53,089,911-53,090,194; hg19). Los *primers* diseñados en ambos casos no amplifican ninguna otra región según las predicciones *in silico* realizadas mediante la herramienta *BLAT* del browser de UCSC. Además, tampoco contienen ningún SNP a lo largo de su secuencia que pueda interferir en el proceso de amplificación mediante PCR.

La PCR fue optimizada y los productos de la reacción se chequearon para comprobar su concentración y pureza a las condiciones dadas. En las condiciones optimizadas de PCR, se obtuvo una única banda a 286 pb correspondiente con el amplicón y con una concentración de 43,27 ng/ μ L.

La región amplificada de UC435 (chr18:53,089,911-53,090,194; hg19) presenta un único dominio de desnaturalización bastante homogéneo según predicción *in silico* de la curva de fluorescencia (a partir de las predicción del programa μ MELT), que implica un intervalo de temperatura de poco más de 5 °C en el cual tiene lugar la completa desnaturalización de la doble hebra (**figura 4.2**).

Figura 4. 2. Perfil de desnaturalización del amplicón de UC435 mediante la predicción *in silico* con el software μ MELT.



Resultados

El estudio de las muestras de ADN amplificadas de 528 esquizofrénicos y 270 controles, analizadas según sus patrones de desnaturalización no revelaron en primera instancia ninguna variante genética en la región estudiada en las condiciones de detección por defecto. No obstante, 9 muestras mostraron perfiles dudosos que, aplicando condiciones menos estrictas de detección para evitar falsos negativos, revelaron la posible existencia de variantes en esas muestras. Las **figuras 4.3 - 4.4** representan las curvas del gradiente de fluorescencia y de perfiles de fluorescencia normalizados para una de las placas donde una variante dudosa fue detectada en esas condiciones.

Figura 4. 3. Perfiles de fluorescencia normalizados en 96 controles con una posible variante detectada (en rojo) empleando condiciones poco estrictas.

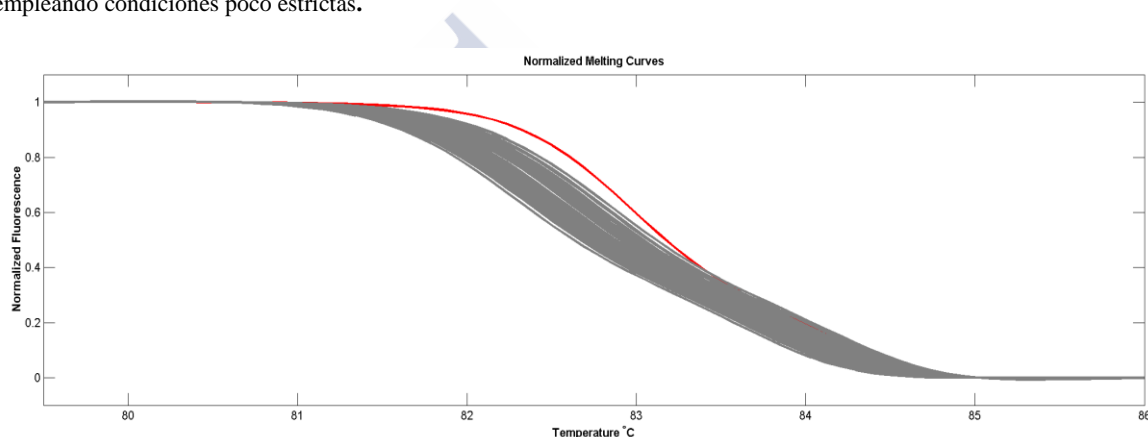
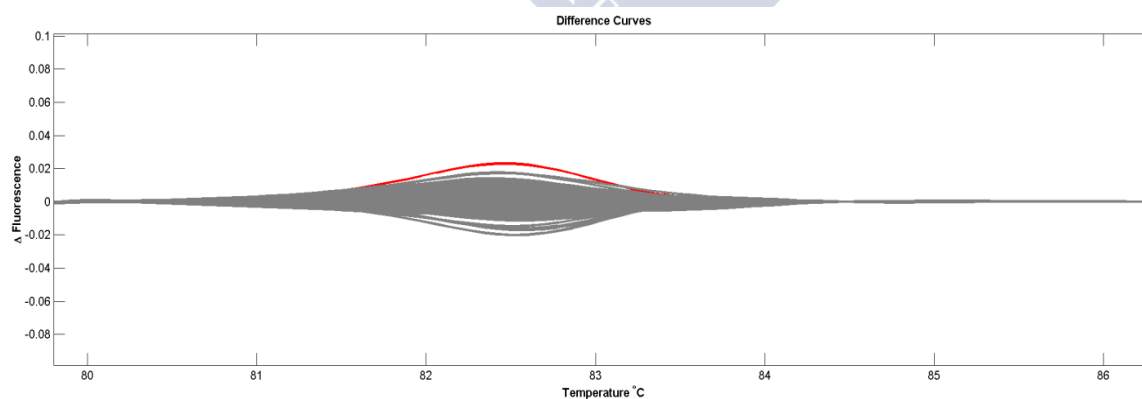


Figura 4. 4. Perfiles de gradiente de fluorescencia en 96 controles sanos con una posible variante detectada (en rojo) empleando condiciones poco estrictas.



I. 4. Secuenciación de las muestras con variantes “dudosas” en UC435

Se optimizó la amplificación para la secuenciación del elemento UCR435 en las muestras con variantes dudosas (5 casos de esquizofrenia y 4 controles). Las 9 muestras y un control negativo se amplificaron siguiendo las condiciones optimizadas, y posteriormente se chequearon mediante un chip de *Agilent 2100 Bioanalyzer*, demostrando que tiene lugar una amplificación específica y con un pico correspondiente al amplicón de UC435 a una concentración de 38,21 ng/μL.

Los resultados de la secuenciación, siguiendo el protocolo de purificación, reacción de secuenciación y posterior análisis mediante la lectura de secuencias no reportaron la detección de ninguna variante en las muestras seleccionadas. De este modo, comprobamos mediante secuenciación de Sanger que las variantes dudosas de HRMA no se correspondían con alteraciones en la secuencia de este elemento ultraconservado en las muestras implicadas.

Mediante esta alternativa económica, 9 muestras mostraron perfiles de desnaturalización de posibles variantes genéticas en la secuencia de UC435, y la secuenciación de las mismas reveló la ausencia de variantes en estas regiones. El análisis de potencia estadística, para la muestra de 528 casos esquizofrénicos, reveló que existe una probabilidad superior al 80% para detectar una variante de frecuencia del 0,15% o mayor, lo cual nos permite descartar la presencia de variantes alélicas en el elemento ultraconservado UC435 de TCF4 con frecuencias similares a las de CNV de riesgo recurrentes en esquizofrenia (Malhotra y Sebat, 2012), que confieren riesgo moderado para la aparición de la enfermedad. Por tanto, son necesarios análisis con tamaños muestrales superiores al de este estudio con el objetivo de estudiar la posible presencia de variantes de menor frecuencia y riesgo mayor, o bien, aunque menos probable, de variantes con efecto más débil.



ESTUDIO II. VARIACIÓN GENÉTICA EN ELEMENTOS ACELERADOS EN HUMANOS EN NPAS3.

II. 1. Introducción

NPAS3 codifica para un factor de transcripción de la familia *bHLH-PAS* (*basic Helix Loop Helix, Period, Aryl hydrocarbon receptor, Single minded*) involucrado en el desarrollo del sistema nervioso central y la neurogénesis en la adultez (Erber-Sieler y cols., 2004; Brunskill y cols., 2005; Pieper y cols., 2005; Wong y cols., 2013). La identificación de una translocación recíproca balanceada que afecta a *NPAS3* en una madre y una hija con esquizofrenia constituye la primera evidencia genética de la implicación de este gen en la enfermedad (Kamnasaran y cols., 2003; Pickard y cols., 2005). Desde entonces, se ha descrito la presencia de variantes raras y comunes asociadas con la susceptibilidad a la esquizofrenia y otros trastornos mentales (Pickard y cols., 2009; MacIntyre y cols. 2010; Yu y cols., 2014). Los estudios de asociación de genoma completo (GWAS) también sugieren un papel de *NPAS3* en la susceptibilidad a los trastornos psiquiátricos, aunque no se ha llegado a alcanzar significación genómica en estas asociaciones (Ferreira y cols., 2008, Huang y cols. 2010; Weber y cols., 2011).

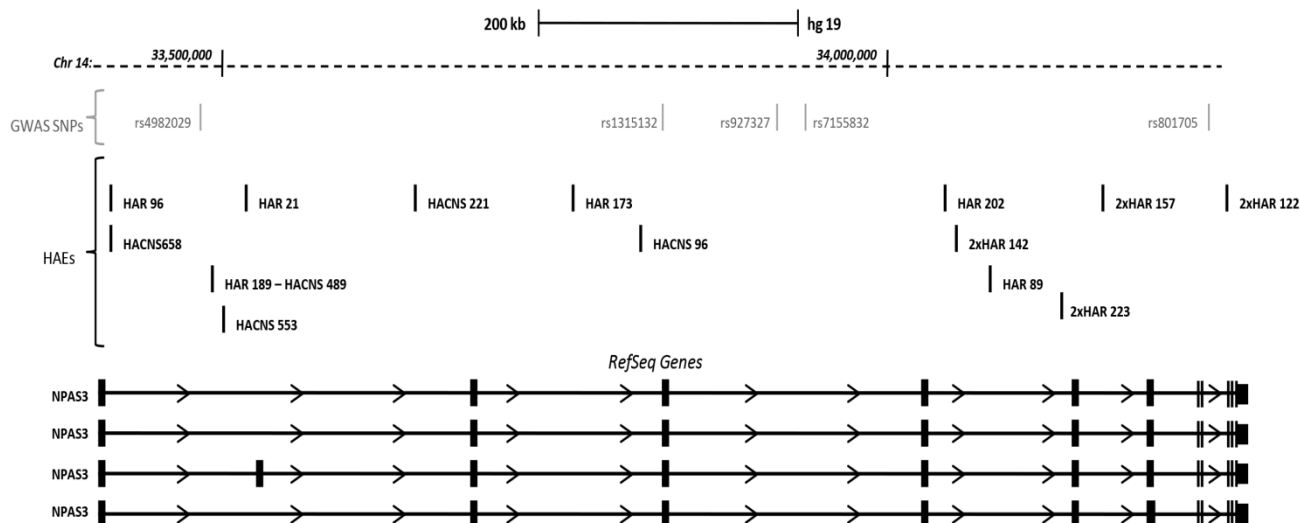
Por otra parte, *NPAS3* es la región genómica con mayor número de elementos reguladores acelerados específicos de humanos (HAEs), es decir, regiones no codificantes altamente conservadas durante la evolución de mamíferos, pero cuya secuencia acumula cambios en el linaje humano tras la escisión respecto al chimpancé (Kamm y cols., 2013a). Hay un total de 14 HAEs en *NPAS3*, y la mayoría de estas regiones funcionan como elementos reguladores de la expresión en el sistema nervioso central en modelos de pez cebra transgénico. Este hecho sugiere un papel crítico de *NPAS3* en la evolución del cerebro humano. Curiosamente, la aceleración es exclusiva de las regiones no codificantes del gen, mientras que la secuencia de codificación está altamente conservada.

Dada la existencia de estas regiones no codificantes de demostrada naturaleza reguladora de la expresión génica, y la haploinsuficiencia de *NPAS3* como factor de riesgo en la aparición de fenotipos relacionados con la esquizofrenia, en este trabajo se postuló que cualquier variante en la secuencia de estos HAEs podría afectar a la susceptibilidad a esquizofrenia. En consecuencia, se realizó el genotipado de los polimorfismos de nucleótido único (SNP) con evidencia de asociación en GWAS de enfermedades psiquiátricas en *NPAS3*, así como de las variantes de nucleótido único raras (SNV) detectadas en proyectos masivos de secuenciación a lo largo de los HAEs de este gen. El análisis de los datos se realizó mediante

Resultados

análisis estadístico diseñado específicamente para el estudio del efecto acumulativo sobre el riesgo de la enfermedad tanto de variantes comunes como raras.

Figura 4. 5. Representación de la estructura de los transcritos de RefSeq del gen *NPAS3*, los 14 HAEs a lo largo del gen y los cinco SNP de estudios GWAS finalmente estudiados.



II.2. Genotipado de variantes en *NPAS3*

Se seleccionaron un total de 25 SNV presentes en población europea de la base de datos de los 1000 genomas a lo largo de los elementos HAEs de *NPAS3*, repartidas en 10 de los 14 HAEs del gen (**figura 4.5**). Además, 5 SNP en *NPAS3* asociados a enfermedades psiquiátricas en estudios de GWAS también fueron incluidas. En total, 30 variantes en *NPAS3* se incluyeron en el diseño de genotipado en Sequenom, pero 4 de ellas se excluyeron por no entrar en el diseño final (**métodos**).

Finalmente, un total de 26 variantes, de las que 21 se encontraban en HAEs de *NPAS3*, así como cinco variantes de GWAS de trastornos psiquiátricos, se genotiparon en 538 casos esquizofrénicos y 539 controles (**tabla 4.1**). Este proceso de genotipado tuvo éxito en más de 99% de las muestras. Asimismo, todas las muestras, menos una, se genotiparon de forma exitosa en más de 95% de las variantes. La totalidad de las 26 variantes analizadas estaban en equilibrio de Hardy-Weinberg ($P > 0,01$), y ocho se encontraban ausentes en las 1077 muestras analizadas. Siete variantes presentan una frecuencia del alelo menor (MAF) inferior al 5%, y tres de ellos una MAF por debajo del 1%.

II. 3. Análisis de variantes en NPAS3

Ninguna de las variantes se asoció con esquizofrenia a nivel individual, como era de esperar teniendo en cuenta nuestro tamaño de la muestra y el bajo efecto de cualquier variante individual en riesgo de la esquizofrenia (**tabla 4.1**).

Tabla 4. 1. Variantes genotipadas en Sequenom para el estudio de la variación genética en HAEs de NPAS3 y de variantes de estudios de GWAS de enfermedades psiquiátricas. Las variantes que se encuentran fuera de los elementos HAEs, en *itálica*, proceden de los siguientes trabajos: rs1315132, $P = 0,0049$ en trastorno por déficit de atención e hiperactividad (Neale y cols., 2010); rs927327, $P = 0,00038$ en esquizofrenia (Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011); rs7155832, $P = 2,27 \times 10^{-5}$ en trastorno bipolar (Sklar y cols., 2011); rs8017055, $P = 0,000851$ en depresión mayor (Ripke y cols., 2013b) y rs4982029, $P = 4 \times 10^{-6}$ en un estudio combinado de trastorno bipolar, depresión mayor y esquizofrenia (Huang y cols., 2010).

SNP ID	HAE	Frec europea 1000G (B)	Alelos A/B	% Genot	Genot SCZ AA/AB/BB	Genot CO AA/AB/BB	valor P HW	valor P alélico
rs10149308	HACNS658	0,001	T/C	100	538/0/0	538/1/0	1	0,318
rs117277815	HACNS96	0,025	T/A	100	527/10/1	519/20/0	0,417	0,160
rs118140829	HACNS658	0,047	C/T	100	501/37/0	488/51/0	0,314	0,137
<i>rs1315132</i>	-	0,139	A/G	99,9	13/12/399	8/115/416	0,566	0,174
rs140407117	HAR21	0,001	G/A	100	538/0/0	539/0/0	1	-
rs141481894	HAR189- HACNS489	0,005	C/T	100	528/10/0	534/5/0	1	0,190
rs143380969	2xHAR157	0,001	A/G	100	536/2/0	536/3/0	1	0,660
rs144275943	HAR89	0,021	T/G	100	515/23/0	518/21/0	1	0,742
rs149570468	2xHAR157	0,005	A/G	100	531/7/0	536/3/0	1	0,200
rs181661714	HAR189- HACNS489	0,001	C/T	100	538/0/0	539/0/0	1	-
rs185406645	HAR202	0,003	A/T	100	538/0/0	539/0/0	1	-
rs189929763	2xHAR223	0,001	T/C	100	538/0/0	539/0/0	1	-
rs190286431	HACNS658	0,001	C/G	100	538/0/0	539/0/0	1	-
rs1958550	HAR89	0,133	T/C	100	411/119/8	399/131/9	0,773	0,375
rs2183276	2xHAR223	0,532	G/A	100	124/259/155	102/271/166	0,827	0,186
rs4627235	2xHAR157	0,259	G/A	99,8	320/184/33	303/202/33	0,591	0,390
<i>rs4982029</i>	-	0,049	G/A	100	519/19/0	510/28/1	0,856	0,118
rs61712153	2xHAR223	0,079	G/A	99,9	460/77/0	465/71/3	0,367	0,952
<i>rs7155832</i>	-	0,789	C/G	99,9	352/155/30	345/170/24	0,061	0,873
rs73264614	HAR96	0,001	T/C	100	538/0/0	539/0/0	1	-
rs75321232	HAR202	0,001	C/T	100	538/0/0	539/0/0	1	-
rs77871170	HAR202	0,007	G/A	100	538/0/0	539/0/0	1	-
rs8003317	HACNS221	0,185	A/G	99,8	348/166/22	360/165/14	0,821	0,370
<i>rs8017055</i>	-	0,266	C/T	99,9	296/191/50	290/204/45	0,021	0,989
rs8017203	HAR96	0,41	T/C	100	193/252/93	208/262/69	0,961	0,093
<i>rs927327</i>	-	0,668	T/C	100	64/238/236	50/236/253	1	0,174

Teniendo en cuenta la evolución excepcional de *NPAS3* (Kamm y cols., 2013a), parece razonable especular que una proporción importante de las variantes tiene un efecto funcional sobre el desarrollo neurológico mediante la alteración de la normal expresión génica de *NPAS3*. Por consiguiente, a pesar de los resultados negativos de las pruebas a nivel individual, se analizaron otros posibles escenarios teniendo en cuenta la totalidad de las variantes en el riesgo asociado, con la ventaja de una mayor potencia estadística en el análisis.

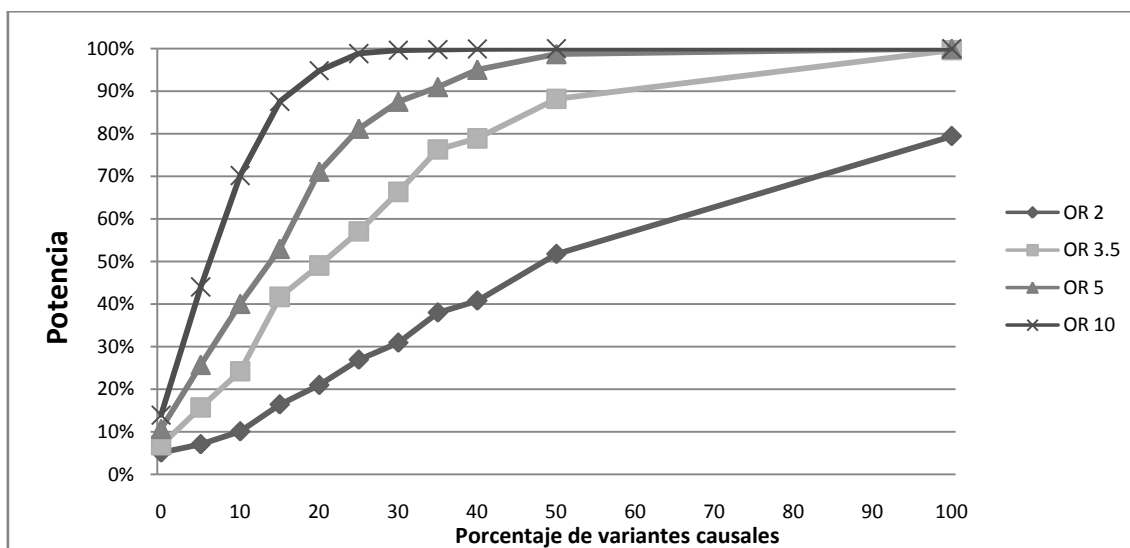
En primer lugar, el test principal realizado probó la acumulación de variantes exclusivamente en los elementos HAEs, y considerando un efecto combinado de las variantes comunes y raras, tanto de riesgo, como de protección. Dado que la prueba no fue significativa ($P = 0,162$), se realizaron pruebas adicionales (un total de 10 con la primera) para explorar diferentes escenarios, desde modelos simples donde sólo variantes raras tienen efecto y siempre de riesgo hasta modelos más complejos donde efectos de riesgo como de protección se tienen en cuenta en variantes raras y comunes, asumiendo diferentes pesos entre ellas en función de la frecuencia. En total, son 5 los escenarios estudiados que, además, se realizaron tanto en variantes presentes en HAEs como en todas las variantes incluyendo las de estudios de GWAS en trastornos psiquiátricos fuera de estos elementos pero a lo largo de *NPAS3*. Ninguno de los modelos probados obtuvo significación positiva (**tabla 4.2**).

Tabla 4. 2. Modelos estadísticos para los 5 diferentes escenarios considerados, incluyendo solo variantes en HAEs o todas las variantes. Los test fueron realizados mediante el paquete estadístico SKAT en R.

Test	variantes raras/comunes	Tipo de test	P (variantes en HAEs)	P (todas variantes)
1	Contribución combinada de variantes raras y comunes	Varianza	0,162	0,21
2	Contribución combinada de variantes raras y comunes	Carga	0,897	0,866
3	Variantes raras y comunes, aumentando el valor de las raras	Combinación de varianza y carga	0,255	0,256
4	Solo contribución de variantes raras	Carga	0,316	0,313
5	Solo contribución de variantes raras	Varianza	0,784	0,813

El análisis de potencia estadística sugiere que el presente trabajo tiene suficiente capacidad para detectar asociación bajo los distintos escenarios realistas probados, de acuerdo con el conocimiento actual de la arquitectura genética de la esquizofrenia (**figura 4.6**). Por ejemplo, existe una potencia estadística del 80% para detectar asociación en un escenario con aproximadamente un 25% de las variantes funcionales y con un OR máximo de 5 para el desarrollo de esquizofrenia.

Figura 4. 6. Potencia estadística para la detección de variantes en función del porcentaje de las variantes analizadas que son causales y el OR máximo asociadas a ellas.



Finalmente, se realizó un análisis de haplotipos de las variantes aquí estudiadas, utilizando un procedimiento de ventanas deslizantes de 2 a 6 SNP (**métodos**). En este último caso, tampoco se halló ninguna asociación significativa tras corregir por las múltiples pruebas realizadas (**tabla 4.3**).

Tabla 4. 3. Análisis de haplotipos mediante procedimiento de ventanas deslizantes de 2 a 6 SNP. Solo los SNP con un valor de frecuencia mínima (MAF) > 0,01 fueron incluidos (a). Los valores p se corrigieron para 1000 permutaciones, lo que lleva a los siguientes valores de p ajustado para los casos de significación nominal (b) de 0,0929, 0,0899, 0,5504, 0,2208 y 0,7922 para las ventanas de 2 a 6, respectivamente.

SNP Ida	Posición (chr14; hg19)	2-SNP	3-SNP	4-SNP	5-SNP	6-SNP
rs8017203	33417163	0,0293	0,0573	0,1434	0,1601	0,0829
rs118140829	33417978	0,0826	0,9566	0,2942	0,2066	0,0009 ^b
rs4982029	33476266	0,291	0,3041	0,1827	0,0148 ^b	0,0602
rs8003317	33649278	0,2401	0,1076	0,0066	0,0776	0,0169
rs117277815	33819429	0,0715	0,0030 ^b	0,0002 ^b	0,1569	0,1536
rs1315132	33829068	0,0232 ^b	0,1935	0,3525	0,2855	0,2018
rs927327	33918155	0,5253	0,7383	0,3272	0,4247	0,4563
rs7155832	33933276	0,8596	0,8591	0,8793	0,5692	0,094
rs1958550	34065521	0,5397	0,6405	0,4371	0,0243	0,0033
rs144275943	34065832	0,5724	0,3153	0,5476	0,5954	
rs61712153	34130195	0,3367	0,6679	0,5119		
rs2183276	34130874	0,4215	0,4399			
rs4627235	34160162	0,4506				
rs8017055	34242983					



ESTUDIO III. RESECUENCIACIÓN DE REGIONES PROMOTORAS DE GENES DE RIESGO EN ESQUIZOFRENIA.

III. 1. Introducción

Recientes hallazgos genómicos sugieren un papel muy relevante para la variación genética en las regiones implicadas en la regulación de la expresión génica en la susceptibilidad a esquizofrenia.

En el caso de los estudios GWAS en esta enfermedad, sólo diez de los 108 *loci* de riesgo identificados en el trabajo más reciente y de mayor volumen en esquizofrenia puede ser atribuible a variantes no sinónimas (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Las variantes asociadas con la esquizofrenia en estudios GWAS están enriquecidas en eQTLs presentes en regiones promotoras y *enhancers* (Richards y cols., 2012; Roussos y cols., 2014).

Por otro lado, la importancia de las variantes en el número de copias (CNV) en el riesgo para la aparición de trastornos del neurodesarrollo también sugiere un papel relevante para la regulación de la expresión génica en el riesgo a esquizofrenia, ya que un mecanismo probable por el cual la CNV ejerce su efecto CNV es el cambio en los niveles de expresión de los genes afectados por la alteración en el número de copias (Henrichsen y cols., 2009). Este mecanismo ha sido demostrado para varios genes de la delección y duplicación en 16p11.2, la delección en 22.q11.2, la duplicación en 15q11.2 o la delección en 3q29 (Mehta y cols., 2014; Migliavacca y cols., 2015; Ye y cols., 2012).

Por último, los estudios de secuenciación del exoma completo en esquizofrenia no lograron encontrar variantes poco comunes de efecto moderado (Need y cols., 2012; Purcell y cols., 2014), lo cual apoya la transición de los estudios genéticos hacia otros centrados en la búsqueda de variantes de susceptibilidad en regiones no codificantes de potencial regulador en los genes implicados en el desarrollo de la enfermedad.

A pesar de las citadas evidencias, hoy en día no hay suficientes estudios sobre la variación genética reguladora en el contexto de la esquizofrenia debido, en mayor medida, a la dificultad en la interpretación de las consecuencias de estas variantes. La variación genética en regiones reguladoras está probablemente más asociada a los cambios específicos en tejidos en donde la expresión del gen o genes regulados sea de mayor importancia, así como a los diferentes estados durante el desarrollo. Por lo tanto, la consideración de la regulación espacial y temporal de la expresión génica puede ser de gran relevancia a la hora de interpretar la variación

genética de potencial regulador. En este contexto, varios proyectos a gran escala surgieron en los últimos años, tales como el Proyecto ENCODE (ENCODE Project Consortium, 2012), cuyo objetivo perseguido fue el de construir un catálogo de variantes funcionales a lo largo del genoma, el Proyecto Roadmap (Kundaje y cols., 2015), que se centró en los cambios de estado de la cromatina durante el desarrollo basado en los patrones de modificación de histonas o BrainSpan (Kang y cols., 2011), cuyo objetivo fue el estudio de los cambios transcripcionales del cerebro humano durante el desarrollo. Además, varias herramientas bioinformáticas surgieron a la luz de los grandes volúmenes de datos para canalizar esa información relativa a las regiones reguladoras presentes en el genoma humano, y que ayudan a la interpretación de la variación genética, tales como RegulomeDB (Boyle y cols., 2012) y AVIA (Vuong y cols., 2012).

En el presente trabajo, se secuenciaron las regiones reguladoras promotoras alrededor de los sitios de inicio de transcripción (TSS) de diferentes genes asociados con la esquizofrenia en estudios GWAS, así como de genes candidatos en CNV de riesgo, para la búsqueda de variantes funcionales raras implicadas en la susceptibilidad a esquizofrenia, y bajo la hipótesis de que la variación rara no es posible capturarla mediante estudios de GWAS. La interpretación de estos resultados se realizó utilizando datos funcionales de grandes proyectos como el proyecto ENCODE, Roadmap, y mediante las predicciones de funcionalidad de RegulomeDB.

III. 2. Cobertura de las regiones analizadas

En este estudio se secuenciaron 30 regiones promotoras de genes de CNV y de GWAS relacionados con esquizofrenia (métodos), comprendiendo un total de 77854 para analizar de forma comparativa la variación genética presente en ellas en 516 casos y 516 controles agrupados en pools de 6 muestras (86 pools de casos y de controles), y a su vez, para el proceso de secuenciación, agrupadas en 4 pools de secuenciación (conteniendo, cada uno, 22 o 21 pools de casos y controles).

Previo al análisis de variantes, se procedió a analizar la cobertura de las regiones secuenciadas. Uno de los amplicones inicialmente amplificados, AMP39 de *GABRA5*, no fue correctamente mapeado y no se pudo analizar, estando el análisis restringido a las 73997 pb restantes, de 29 regiones promotoras.

En ellas, el análisis de cobertura comparado en casos y controles reveló que no existen diferencias de cobertura significativas entre casos y controles (**figura 4.7**). De hecho, existe, de forma no significativa, una mayor cobertura en el conjunto de controles analizados que en los

casos (cobertura en controles = $0,90 \pm 0,06$, cobertura en casos = $0,89 \pm 0,06$; valor P de la prueba de Kruskal-Wallis = 0,122).

Figura 4. 7. Diagrama de cajas y bigotes para la comparación de los valores de cobertura de las regiones analizadas en casos frente a controles.

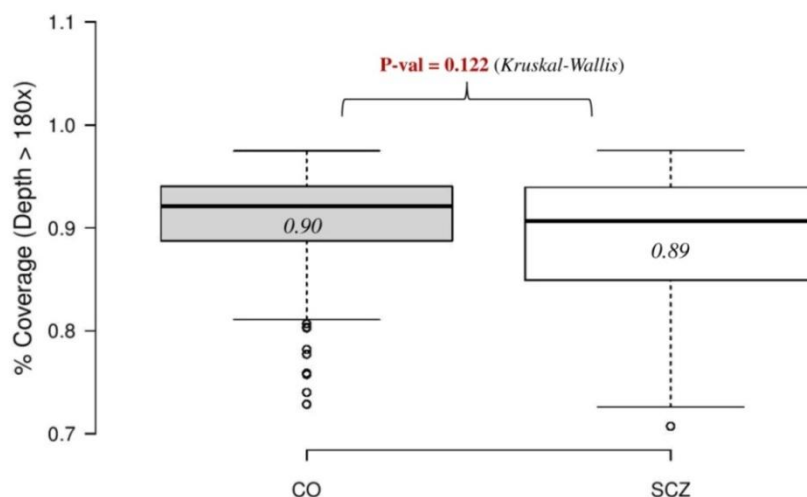
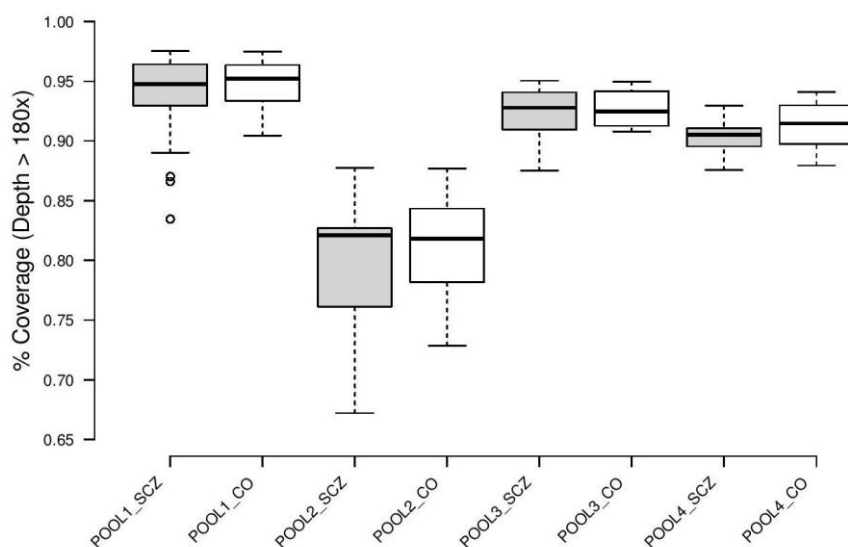


Figura 4. 8. Diagrama de cajas y bigotes para la comparación de los valores de cobertura de las regiones analizadas en casos frente a controles para cada uno de los 4 pools de secuenciación.



Sin embargo, sí existe diferencia en la cobertura existente en los 4 pools de secuenciación, tal y como se puede apreciar en la **figura 4.8**. Para evitar la infrarrepresentación de los individuos pertenecientes al pool 2 de secuenciación, se restringió el análisis a aquellas posiciones genómicas con una profundidad de lectura mínima de 180 en el 95% de los

individuos (**tabla 4.4**). Un total de 55729 posiciones fueron consideradas mediante este filtro (que representan un 75% del total de posiciones analizadas). Más de un 70% de las 55729 posiciones analizadas tuvieron, además, una DP > 240 en un 95% de los individuos seleccionados (**tabla 4.5**).

Tabla 4. 4. Cobertura de las regiones secuenciadas a DP > 180 en función del número de individuos cubiertos

Porcentaje de pools (172) cubiertos	Nº posiciones (pb)	% sobre el total de posiciones
100%	28243	38.17%
99%	46561	62.92%
95%	55729	75.31%
90%	58538	79.11%
80%	28243	84.30%

Tabla 4. 5. Cobertura de las regiones secuenciadas a DP > 240 en función del número de individuos cubiertos

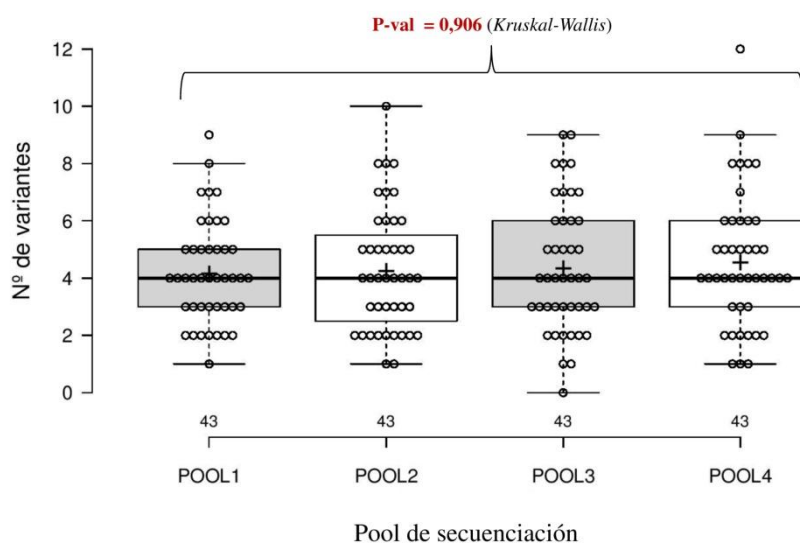
Porcentaje de pools (172) cubiertos	Nº posiciones (pb)	% sobre el total de posiciones
100%	22848	30.88%
99%	40280	54.43%
95%	51856	70.08%
90%	54500	73.65%
80%	58976	79.70%

III. 3. Detección de variantes

Con los filtros de calidad descritos (métodos), y el filtro posicional basado en una profundidad de lectura superior a 180 en el 95% de las muestras analizadas, se detectaron un total de 297 posiciones con variante genética rara (a frecuencias inferiores a 1% en las bases de datos y en el propio conjunto de muestras). 361 variantes en casos y 388 en controles fueron el total de SNV e indels de frecuencia rara presentes en el presente estudio.

Ninguno de los pools de secuenciación se halló enriquecido en las 749 variantes detectadas tras los filtros pertinentes (**figura 4.9**), a pesar de las diferencias de cobertura observadas en una primera instancia (**figura 4.8**), lo que demuestra que el filtro posicional aplicado de eliminar aquellas variantes presentes en posiciones altamente cubiertas (DP > 180) en menos del 95% de las muestras fue exitoso a la hora de eliminar posibles sesgos entre los 4 pools de secuenciación.

Figura 4. 9. Diagrama de cajas y bigotes para la comparación entre el número de variantes detectadas en cada pool de muestras en los distintos pools de secuenciación.



Además, el análisis de las frecuencias internas, dentro de cada pool, de las variantes detectadas sugirió que la enorme mayoría de las mismas es real, y que los filtros aplicados, si bien hubieron podido eliminar algunas variantes verdaderas (falsos negativos), consiguieron eliminar la mayoría de los artefactos del proceso de secuenciación (falsos positivos). Para un total de 12 alelos presentes en cada pool, 2 por cada individuo representado en ellos, en condiciones de equimolaridad entre las muestras integrantes, se esperaría que una variante rara presente en un alelo de un pool tuviera una frecuencia aproximada de 0.08, siendo este caso el más frecuente. La distribución de frecuencias de las 749 variantes detectadas revela un comportamiento similar (**figura 4.10**), donde la mediana de la misma presenta un valor de 7,4.

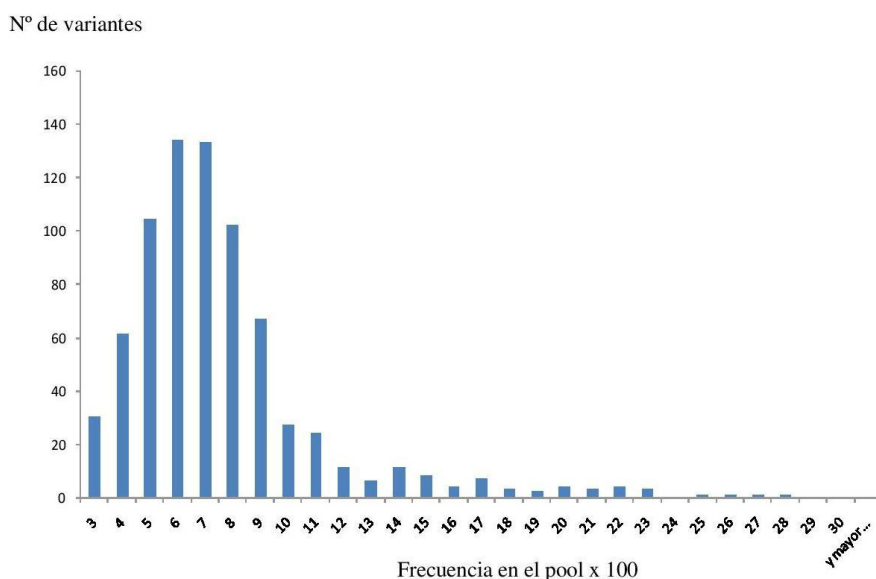
III. 4. Confirmación de variantes secuenciadas mediante genotipado en Sequenom

De las 297 posiciones con 749 variantes en las muestras analizadas, las 56 posiciones con predicción de funcionalidad "2" en RegulomeDB (con un total de 106 variantes) fueron escogidas para su confirmación mediante genotipado en la plataforma de Sequenom (**ANEXO IV**). Este subconjunto fue seleccionado ya que, si bien la probabilidad de falsos positivos esperada es la misma en cualquier grupo de variantes, la probabilidad de verdaderos positivos existentes en esta categoría será la menor, dada su mayor funcionalidad y las consecuencias más relevantes que acarrearían estas variantes. Por tanto, la tasa de falsos positivos será mayor para estas posiciones

De las 106 variantes en RegulomeDB 2, 89 entraron finalmente en los diseños multiplex de Sequenom, por lo que la tasa de falsos positivos se calculó a partir de este conjunto

(ANEXO V). 86 de esas 89 variantes fueron confirmadas, lo que revela una tasa de falsos positivos en el conjunto más propenso a tenerlos de 3,4%, correspondientes a 2 variantes en controles y una en casos. Dada esta baja tasa de errores, que demuestra la efectividad de los filtros seleccionados en la eliminación de variantes no reales, estas 3 fueron excluidas y el análisis, a partir de entonces, se realizó sobre las 360 y 386 variantes en casos y controles, respectivamente, en 294 posiciones afectadas.

Figura 4. 10. Distribución de la frecuencia dentro de cada pool de las variantes presentes en el conjunto final de 749.



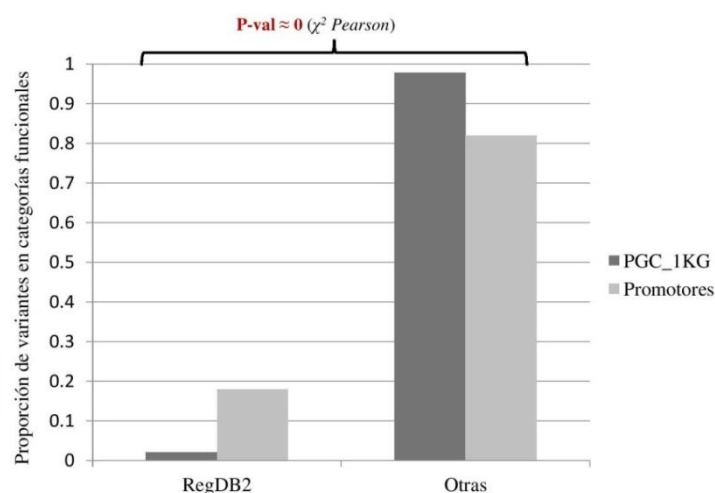
III.5. Comparaciones caso-control basadas en RegulomeDB

En primer lugar, se analizó la incidencia de las variantes detectadas y filtradas según lo explicado en la categoría de RegulomeDB de mayor consecuencia funcional (categoría 2). Un total de 103 variantes de las 746 aquí detectadas, tras eliminar los falsos positivos de secuenciación, pertenecen a esta categoría funcional (un 13,8%), lo cual ya revela, en lo relativo a estas regiones, una selección funcional muy significativa en comparación con el resto del genoma (la tasa esperada en la totalidad del genoma es del 1,48%)

A modo de comparación, se seleccionaron las variantes de dbSNP v.141, anotadas en RegulomeDB, genotipadas por el PGC (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) y no presentes en alguna de las poblaciones secuenciadas en la fase 3 del proyecto de los 1000 genomas a una frecuencia superior al 1%. Un total de 4475 posiciones con variantes tuvieron categoría funcional RegulomeDB 2 de un total de 212681

seleccionadas de este modo (2,1%). En comparación, en las variantes de este estudio, 53 posiciones presentan categoría funcional 2 del total de 294 posiciones mutadas (18%; valor P de la prueba de χ^2 de Pearson ≈ 0), lo que demuestra, nuevamente, que la selección de 73997 pb finalmente mapeadas están fuertemente enriquecidas en regiones de gran relevancia para la regulación de la expresión génica (**figura 4.11**).

Figura 4. 11. Comparación entre el número de posiciones genómicas con variante en RegulomeDB 2 entre las presentes en las regiones secuenciadas y aquellas presentes en las variantes genotipadas por el PGC, presentes en dbSNP v.141 y a frecuencia menor al 1% en todas las poblaciones de los 1000 genomas (PGC_1KG)



Respecto a la comparación entre casos y controles, existe un enriquecimiento estadísticamente significativo en las variantes RegulomeDB 2 presentes en casos respecto a controles (**figura 4.12**; OR = 1,92 (1,26 - 2,95); P de la prueba exacta de Fisher de dos colas = 0,0028). Esto revela que las variantes genéticas reguladoras en las regiones promotoras de los genes seleccionados son un factor de riesgo en el desarrollo de esquizofrenia.

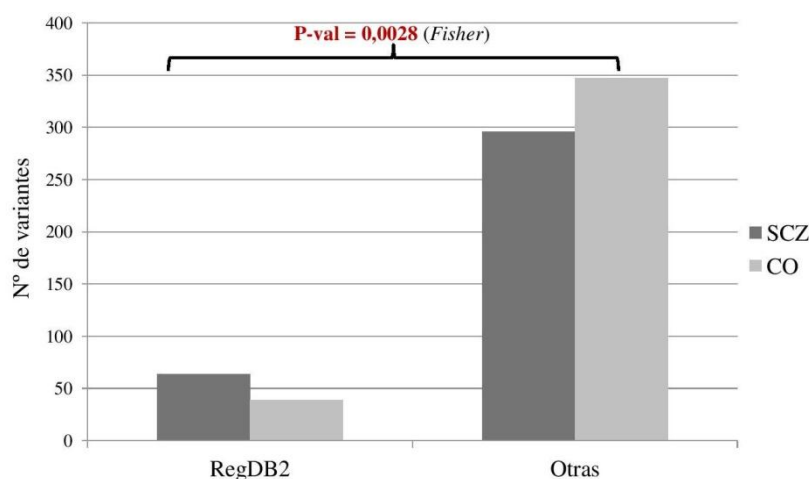
III.6. Comparaciones caso-control basadas en los estados de la cromatina (Proyecto Roadmap)

La variación genética en regiones reguladoras está probablemente más asociada a cambios específicos en aquellos tejidos donde la alteración expresión del gen o genes estudiados durante las etapas del desarrollo sea más relevante para la aparición del fenotipo esquizofrénico. En este sentido, para profundizar en la funcionalidad de la variación genética presente en las regiones promotoras secuenciadas, se procedió al análisis del impacto de la misma en los

Resultados

estados funcionales de esas regiones en los tejidos más implicados en el desarrollo de la esquizofrenia.

Figura 4. 12. Enriquecimiento de variantes en RegulomeDB 2 en casos respecto a controles en las regiones promotoras secuenciadas. En la figura se representa el número de variantes en la categoría 2 de RegulomeDB frente al resto de categorías funcionales.



Haciendo uso de los datos del reciente proyecto Roadmap, se analizó la incidencia de variantes genéticas en los distintos estados de la cromatina (descritos en función de modificaciones epigenéticas; **tabla 1.9**) de los tejidos más relevantes en esquizofrenia en regulación genética según estudios GWAS (Giro angular y corteza prefrontal medial) y en cerebro fetal, dada la presencia, en este estudio, de genes de CNV implicadas en trastornos del neurodesarrollo (**tabla 4.6**).

En este análisis se detectaron diferencias significativas en la incidencia de variantes en casos y controles, a nivel nominal de 0,05, en el estado represor Polycomb a lo largo de todos los tejidos analizados menos en cerebro fetal femenino. Igualmente, otro estado parcialmente reprimido, *enhancer* bivalente, también tuvo diferencias significativas en el número de variantes en casos y controles en cerebro fetal masculino, estando sobrerrepresentado en casos en la totalidad de los tejidos. Existe, además, una fuerte correlación en los estados de la cromatina de los tejidos estudiados, y la mayoría de las variantes situadas en estados represores Polycomb en algún tejido también lo son en el resto, o en su lugar, *enhancer* bivalente, que está funcionalmente relacionado. Por otro lado, no se observaron diferencias apreciables en los estados activos de la cromatina en los tejidos analizados, lo que sugiere que en las regiones promotoras existe una mayor relevancia de las funciones represivas o bivalentes, en contra de lo observado en los *loci* asociados a esquizofrenia en el GWAS del PGC (Schizophrenia Working

Group of the Psychiatric Genomics Consortium, 2014), que se encontraban sobrerrepresentados en enhancers activos en los tejidos aquí analizados, y haciendo uso de los mismos datos de Roadmap.

Además, 9 de las 360 variantes en casos frente a ninguna en las 386 de controles se encuentran situadas en posiciones de estados de represión Polycomb y RegulomeDB 2, una diferencia altamente significativa (P valor de prueba exacta de Fisher de dos colas = 0,0011, OR = 20,89, 95% C.I. = 2,64 - 2696,93). Esta diferencia también es altamente significativa en la comparación de casos portadores y controles portadores de variantes (P valor de prueba exacta de Fisher de dos colas= 0,0016, OR = 19,34, 95% C.I. = 2,45-2495,26).

Estas 9 variantes de riesgo detectadas se encuentran, además, únicamente en los promotores de 2 genes (*HNFI1B* y *LHX1*) pertenecientes a la misma CNV (17q12) (**tabla 4.7**). De hecho, 2 de esas variantes se encuentran tan solo separadas 2 pb, estando todas ellas presentes en individuos diferentes.

De acuerdo con la clasificación de los estados de la cromatina de Roadmap basados en marca de histonas, todas menos una de estas SNV se ubican dentro de un pico ChIP-seq para *EZH2*, un miembro del complejo Polycomb represivo 2 (PRC2) en diferentes líneas celulares de los datos de ENCODE. Además, una de estas SNV se encuentra en un pico ChIP-seq para *SUZ12*, otro componente del PRC2, y *REST*, un factor de transcripción supuestamente implicado en el reclutamiento del complejo represivo PRC2. Por otro lado, la única SNV presente en más de un caso cae dentro de un pico ChIP-seq para *YY1*, otro factor de transcripción implicado en el reclutamiento del complejo PRC2 (Lanzuolo y cols., 2012; Margueron y cols., 2011) Estos datos evidencian la importancia de estas variantes en la correcta represión de la expresión génica de los genes *HNFI1B* y *LHX1* de la CNV 17q12.

Tabla 4. 6. Enriquecimiento de las variantes detectadas en casos frente a controles en los 15 estados de la cromatina (Proyecto Roadmap) de los siguientes tejidos: giro angular, corteza prefrontal medial y cerebro fetal masculino y femenino. Para cada tejido se representa el número de variantes que caen en casos (SCZ) y en controles (CO) en los estados correspondientes. Las pruebas de enriquecimiento se realizaron mediante test exacto de Fisher de dos colas.

Estado cromatina (Roadmap)	Descripción	Giro Angular		Corteza prefrontal medial		Cerebro fetal (hombre)		Cerebro fetal (mujer)	
		SCZ/CO	P Fisher (2 colas)	SCZ/CO	P Fisher (2 colas)	SCZ/CO	P Fisher (2 colas)	SCZ/CO	P Fisher (2 colas)
1_TssA	TSS activo	144/175	0,159	142/158	0,709	97/116	0,373	143/165	0,414
2_TssAFlnk	Flancos de TSS activo	57/57	0,76	38/43	0,815	25/40	0,119	32/33	0,897
2_TxFlnk	Transcrito en 5' y 3'	0/0	-	0/0	-	0/0	-	0/0	-
4_Tx	Elevada transcripción	10/15	0,424	22/27	0,66	9/13	0,523	22/32	0,262
5_TxWk	Débil transcripción	40/41	0,906	58/58	0,687	59/64	1	53/46	0,281
6_EnhG	Enhancers en gen	0/3	0,25	0/4	0,125	0/0	-	8/8	1
7_Enh	Enhancers	53/50	0,525	24/20	0,438	35/47	0,294	4/5	1
8_ZNF/Rpis	Genes de ZNF y repeticiones	0/0	-	0/0	-	14/7	0,12	0/0	-
9_Het	Heterocromatina	0/0	-	0/0	-	0/0	-	0/1	1
10_TssBiv	TSS Bivalente	0/0	-	18/29	0,176	7/6	0,783	12/7	0,246
11_BivFlnk	Flanco de TSS/enhancer bivalente	2/2	1	1/0	0,483	10/8	0,635	8/6	0,594
12_EnhBiv	Enhancer bivalente	3/1	0,357	8/3	0,132	18/8	0,044	9/3	0,081
13_ReprPC	Represor Polycomb	17/5	0,008	14/5	0,034	12/4	0,041	11/4	0,066
14_ReprPCWk	Represor Polycomb débil	0/0	-	19/17	0,611	0/0	-	0/0	-
15_Quies	Quiescente	34/37	1	16/22	0,506	74/73	0,582	52/60	0,683

Tabla 4. 7. Datos de ChIP-seq para las líneas celulares analizadas en el proyecto Encode y estados de la cromatina del proyecto Roadmap para los tejidos analizados en el presente estudio para las nueve SNV presentes en estado represor Polycomb en cualquiera de los tejidos seleccionados.

Chr	Posicion	Gen (promotor)	Distancia a TSS (bp)	N _{roots} SCZ	N _{roots} CO	Regulome DB	Señal en ChIP-Seq (RegulomeDB)	Giro angular	Corteza prefrontal medial	Cerebro fetal (hombre)	Cerebro fetal (mujer)
chr17	35293056	LHX1	1028	1	0	2b	EZH2	13_ReprPC	12_EnhBiv	12_EnhBiv	12_EnhBiv
chr17	35293058	LHX1	1026	1	0	2b	EZH2	13_ReprPC	12_EnhBiv	12_EnhBiv	12_EnhBiv
chr17	35293268	LHX1	816	1	0	2b	EZH2	12_EnhBiv	13_ReprPC	12_EnhBiv	11_BivFlnk
chr17	36104916	HNFB1B	-93	1	0	2b	EZH2, SUZ12, REST	13_ReprPC	13_ReprPC	13_ReprPC	13_ReprPC
chr17	36105539	HNFB1B	302	4	0	2b	EZH2, YY1	13_ReprPC	13_ReprPC	13_ReprPC	13_ReprPC
chr17	36106342	HNFB1B	1105	1	0	2b	-	13_ReprPC	13_ReprPC	13_ReprPC	13_ReprPC





ESTUDIO IV. ANÁLISIS DE TRANSCRIPTOMA DE GIRO CINGULADO EN ESQUIZOFRENIA. ESTUDIO DE VARIACIÓN GENÉTICA Y DIFERENCIAS DE EXPRESIÓN EN CEREBRO

IV. 1. Introducción

El giro cingulado es una parte integral del sistema límbico que se localiza en la circunvolución o *gyrus* del área media del cerebro, y que se haya involucrado en la formación y procesamiento de la emoción (Hadland y cols., 2003), aprendizaje y memoria (Sutherland y cols., 1988). Es destacable su importancia en los trastornos del estado de ánimo (Drevets y cols., 2008). Las anomalías en el volumen de esta región cerebral (Costain y cols., 2010; Takahashi y cols., 2003), en el metabolismo (Haznedar y cols., 2004), en la conectividad (Wang y cols., 2015) y en la expresión de genes de astrocitos (Katsel y cols., 2011), entre otras, se han relacionado, además, en numerosas ocasiones con la patofisiología de la esquizofrenia. Por estas razones, el análisis de los genes diferencialmente expresados en esquizofrenia en esta región cerebral es de gran relevancia para el estudio de la regulación de la expresión en el contexto de la esquizofrenia y otras enfermedades del neurodesarrollo relacionadas.

Una de las grandes ventajas de las tecnologías de ultrasecuenciación ha sido su aplicación a los estudios de expresión, donde la técnica de RNA-seq representa una notable mejoría frente a sus predecesoras de microarrays (**apartado 1.3.2**). La abundancia del transcrito es directamente proporcional al número de lecturas de secuenciación que se mapean, dando como resultado tasas menores de falsos negativos y falsos positivos, además de un mayor rango lineal de detección (Mortazavi y cols., 2008; Wilhelm y cols., 2008).

Por otro lado, mediante las tecnologías de ultrasecuenciación se ha permitido el estudio del exoma en esquizofrenia en profundidad (Fromer y cols., 2014; Purcell y cols., 2014), lo que ha llevado al conocimiento de grupos de genes implicados en la etiología de la enfermedad, como aquellos relativos a los procesos sinápticos, y posiciones de relevancia para la regulación post transcripcional, como sitios de unión de la proteína FMRP (Fromer y cols., 2014; Purcell y cols., 2014) o de microRNA (Warnica y cols., 2015). Sin embargo, son menores los estudios que han aplicado la tecnología de ultrasecuenciación para la detección de variación genética en ARN de tejido específicamente relevante en esquizofrenia, lo que permite conocer las variantes genéticas presentes en genes expresados en esos tejidos. Teniendo en cuenta las numerosas evidencias y mayor relevancia que están adquiriendo los mecanismos de regulación post-transcripcional en la etiología de las enfermedades psiquiátricas, el estudio de la variación genética diferencial entre casos y controles de forma conjunta con las diferencias de expresión

observadas en los genes analizados representa una estrategia muy potente para el análisis *in situ* del efecto de las variantes genéticas en el transcriptoma de cerebro en esquizofrenia, algo posible gracias al desarrollo de la tecnología de RNA-seq.

El análisis de la coexpresión de los distintos genes en cerebro a lo largo del desarrollo refleja una muy clara relación con los principales tipos celulares como neuronas, oligodendrocitos, astrocitos o microglia, por lo que el estudio de los datos de expresión en comparación con los patrones de coexpresión en cerebro es una alternativa muy potente al estudio agnóstico de expresión sin hipótesis previas.

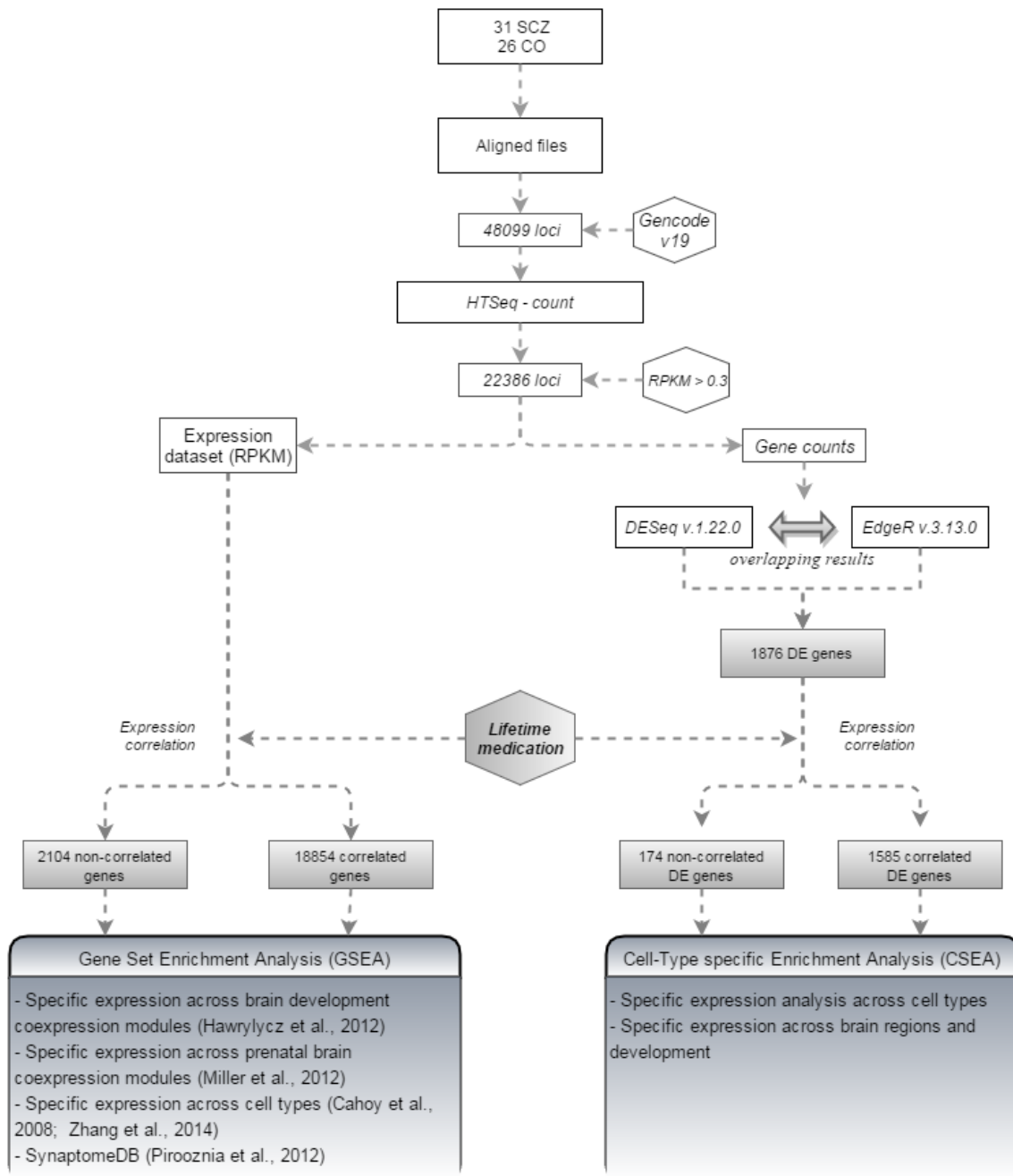
Uno de las mayores dificultades, sin embargo, de los estudios del transcriptoma en esquizofrenia es que la mayoría de los pacientes toman medicamentos antipsicóticos a lo largo de su vida, lo que puede alterar el patrón de expresión génica en enfermos (Crespo-Facorro y cols, 2015), al igual que sucede con el alcohol u otras sustancias de abuso (Liu y cols., 2006a; Torres y Horowitz, 1999). Por ello, es importante el desarrollo de estrategias que permitan eliminar la influencia de estos factores para poder evaluar las diferencias de expresión entre ambos grupos.

En este sentido, en el presente trabajo se analizaron comparativamente los transcriptomas de giro cingulado en muestras *post mortem* bien caracterizadas de individuos esquizofrénicos y controles, siguiendo una estrategia dirigida a la comparación con módulos de genes coexpresados en cerebro. Se estableció, en el mismo, un análisis en paralelo de la expresión génica en el tejido bajo estudio, teniendo en cuenta la dependencia con las posibles variables condicionantes como la medicación administrada, y un análisis de las variantes en aquellos genes consistentemente expresados en el total de los individuos analizados (en los genes más representados en las muestras de cerebro). Finalmente, se estudió la posible conexión entre las variantes presentes en los genes expresados y los cambios en los niveles de expresión génica de los mismos. Mediante el análisis de la relación entre la distribución de variantes en las diferentes categorías de genes presentes en el transcriptoma cerebral, el presente estudio ha permitido identificar genes candidatos que podrían ofrecer nuevas e importantes ideas en relación al riesgo subyacente al desarrollo de la esquizofrenia por cambios en la expresión génica.

IV. 2. Estudio de expresión en giro cingulado en esquizofrenia

En el presente trabajo, en primer lugar, se analizó la expresión génica a partir de muestras de ARN *post mortem* de giro cingulado de 31 casos esquizofrénicos y 26 controles sanos, siguiendo el protocolo de RNA-seq descrito previamente (**métodos, figura 4.13**).

Figura 4. 13. Procedimiento de análisis en el estudio de la expresión diferencial en esquizofrenia, y estrategias para la interpretación de las diferencias de expresión mediante GSEA y CSEA.



IV.2.1. Análisis de genes con expresión detectable

A partir de la base de datos de Gencode v.19 comprehensive, se definieron 48099 *loci* no solapantes en los cuales se estudió la expresión en giro cingulado de pacientes esquizofrénicos y controles sanos. 22386 de los 48099 *loci* se hallaron por encima de los valores de expresión definidos como detectables (RPKM > 0,3; **métodos**).

Dada la menor expresión de los genes no codificantes frente a los que sí lo son, tal y como se describió a lo largo del proyecto ENCODE (Djebali y cols. 2012), se estudió la naturaleza de aquellos que se encontraban expresados de forma inequívoca en las muestras de cerebro analizadas (**tabla 4.8**).

En coherencia con los resultados de ENCODE, los genes con expresión detectable se encuentran enriquecidos en aquellos codificantes de proteínas en comparación con el total de *loci* analizados (60,1% frente a 36,8 %, valor P de la prueba χ^2 de Pearson ≈ 0). No obstante, 8935 genes no codificantes también se encontraron expresados de forma consistente en las muestras analizadas, lo que revela la importancia de éstos en la naturaleza del transcriptoma cerebral.

Tabla 4. 8. Naturaleza de los genes expresados en giro cingulado en el presente estudio. En la tabla se representa el número de *loci* que son genes codificantes o contienen al menos algún gen codificante de todos los *loci* analizados o de aquellos con expresión detectable.

	N <i>loci</i>	<i>loci</i> codificantes	<i>loci</i> no codificantes
Todos los genes	48099	17692 (36,8%)	30407
Exp. detectable (RPKM > 0,3)	22386	13451 (60,1 %)	8935

IV.2.2. Estudio de correlación con medicación antipsicótica

Dada la influencia demostrada de la medicación a lo largo de la vida en la expresión génica (Crespo-Facorro y cols, 2015), se utilizaron los valores P corregidos (BY-p) de las correlaciones de Pearson calculadas entre los niveles de expresión génica para cada uno de los 22386 *loci* expresados y la dosis de medicación a lo largo de la vida de los individuos cuyas muestras se utilizan en este trabajo (**métodos**). En base a estas correlaciones, los *loci* expresados en cerebro se subdividieron en 3 grupos que se analizarán comparativamente a lo largo del presente trabajo:

- Todos los genes con expresión detectable (22386 *loci* con RPKM > 0,3)
- Genes no correlacionados con la medicación a lo largo de la vida (2104 *loci* con valores P de las correlaciones de Pearson corregidos BY-p > 0,05)
- Genes fuertemente correlacionados con la medicación a lo largo de la vida (18854 *loci* con valores P de las correlaciones de Pearson corregidos BY-p < 5×10^{-5}), de los cuales 10417 (55,3%) y 8437 (44,7%) están directamente o inversamente correlacionados, respectivamente.

Tabla 4. 9. Naturaleza de los genes correlacionados y no correlacionados con la medicación antipsicótica a lo largo de la vida. En la tabla se representa el número de *loci* que son genes codificantes o contienen al menos algún gen codificante de todos los *loci* analizados o de aquellos con expresión detectable.

Correlación con medicación	N <i>loci</i>	<i>loci</i> codificantes	<i>loci</i> no codificantes	% <i>loci</i> codificantes
No correlacionados	2104	1295	809	61,5%
Directamente correlacionados	10417	6206	4211	59,5%
Inversamente correlacionados	8437	5060	3377	59,9%
Todos los genes (exp. detectable)	22386	13451	8935	60,1%

Tal y como se describe en la **tabla 4.9**, existe una proporción similar de genes de naturaleza no codificante en aquellos no correlacionados y los que sí los están, tanto de forma inversa como directa, con la medicación a lo largo de la vida de los individuos analizados.

IV.2.3. Estudio de enriquecimiento de los datos de expresión en grupos de genes mediante GSEA.

IV.2.3.1. Análisis de enriquecimiento en módulos de coexpresión en esquizofrenia

Para la interpretación biológica de las diferencias de expresión observadas entre casos afectos y controles sanos, se exploró la relación entre los genes transcritos en giro cingulado en nuestro estudio y módulos de coexpresión espaciotemporales descritos anteriormente en cerebro humano (Hawrylycz y cols., 2012; Miller y cols., 2014). Para ello, mediante la herramienta GSEA (del inglés *Gene Set Enrichment Analysis*) (Subramanian y cols., 2005) se estudió la relación entre las diferencias de expresión en casos y controles en el conjunto de genes con expresión detectable (22386 *loci*) y, por un lado, 13 grupos o módulos de genes coexpresados en cerebro a lo largo de la vida (Hawrylycz y cols., 2012) y, por otro, 42 módulos de coexpresión en transcriptoma prenatal (Miller y cols., 2014).

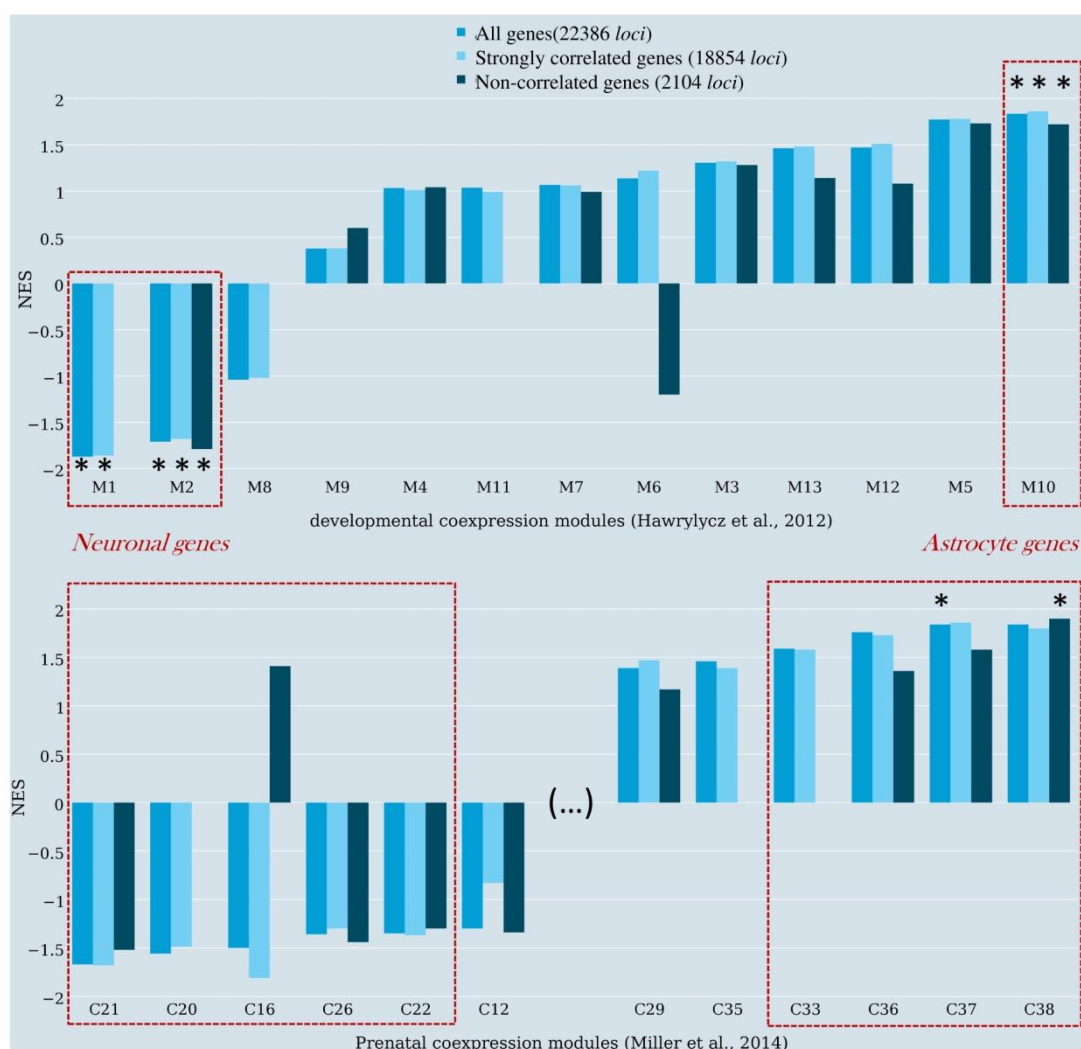
Por lo tanto, teniendo en cuenta la naturaleza de la esquizofrenia como trastorno del neurodesarrollo (Najas-García y cols, 2014, Rapoport y cols, 2012), se probó el enriquecimiento de los módulos señalados en los datos de expresión de todos los genes con expresión detectable (N = 22386 *loci*), sólo los genes no correlacionados con medicación (N = 2104 *loci*) y genes fuertemente correlacionados con la medicación (N = 18854 *loci*).

Se analizó la significación del enriquecimiento en estos tres conjuntos (**figura 4.14, ANEXOVI-VII**) y los niveles de enriquecimiento normalizado (NES) en cada caso. El módulo M2 (Hawrylycz y cols, 2012), compuesto por genes expresados principalmente en neuronas, se halló subexpresado en las muestras de esquizofrenia a través de los tres conjuntos de datos probados, incluso cuando sólo se consideran aquellos genes no correlacionados con los niveles

Resultados

de medicación a lo largo de la vida. Este hecho revela que los resultados no se deben principalmente a efectos de la medicación ($N = 389$ genes, $NES = -1,79$, $FRD-q = 0,02$). El módulo M2 se encuentra fuertemente relacionado con funciones de estrés oxidativo y enriquecido en proteínas de mitocondria.

Figura 4. 14. Resultados del análisis mediante GSEA del enriquecimiento de módulos de coexpresión en genes expresados ($RPKM > 0,3$). En la figura se representa el valor de enriquecimiento normalizado (NES), ordenado de menor a mayor (de izquierda a derecha), de los módulos de coexpresión (Hawrylycz y cols., 2012; Miller y cols., 2014) en los datos de expresión de la totalidad de los genes (22386), de genes fuertemente correlacionados (15854) y de genes no correlacionados con la medicación (2104). Los módulos estadísticamente significativos tras correcciones múltiples (FDR-q) están marcados con un asterisco *. Los valores de enriquecimiento y valores p en cada caso se encuentran descritos en ANEXO VI-VII. Las gráficas de enriquecimiento de GSEA se muestran en ANEXO X-XI.



El módulo M1 (Hawrylycz y cols., 2012), otro módulo de expresión característico de neuronas, también se encontró subexpresado en esquizofrenia, aunque esta vez no pudo ser

probado en los datos de expresión de genes no correlacionados debido a que menos de 10 genes de este módulo solapaban con el conjunto de genes no correlacionados con la medicación, y fueron filtrados (**métodos, figura 4.14**). Sin embargo, fue significativo en el total de genes ($N = 84$ genes, $NES = -1,85$, $FRD-q = 0,01$).

Por otro lado, el módulo de coexpresión M10 (Hawrylycz y cols., 2012), compuesto principalmente por genes característicos de astrocitos, se encontró consistentemente sobreexpresado en las muestras de esquizofrenia a través de los tres grupos de genes estudiados, siendo significativo incluso en aquellos genes independientes de la medicación ($N = 28$ genes, $NES = 1,72$, $FRD-q = 0,05$; **figura 4.14**).

Además, en el análisis de los módulos de coexpresión en cerebro prenatal (Miller y cols., 2014), tres de ellos (C36, C37 y C38), enriquecidos también en genes de astrocitos, fueron los módulos más sobreexpresados en las muestras de esquizofrenia (**figura 4.14**). Uno de ellos, el módulo C38, compuesto mayoritariamente por genes cuya expresión se reduce notablemente con la edad y relacionados con los procesos de replicación, ciclo celular, proliferación celular y ensamblaje de la cromatina, entre otros, así como enriquecido en proteínas de la densidad postsináptica, presentó un enriquecimiento estadísticamente significativo en los datos de expresión de genes no correlacionados con la medicación ($N = 37$ genes, $NES = 1,9$, $FRD-q = 0,007$), siendo el único de los módulos de Miller y cols en presentar significación.

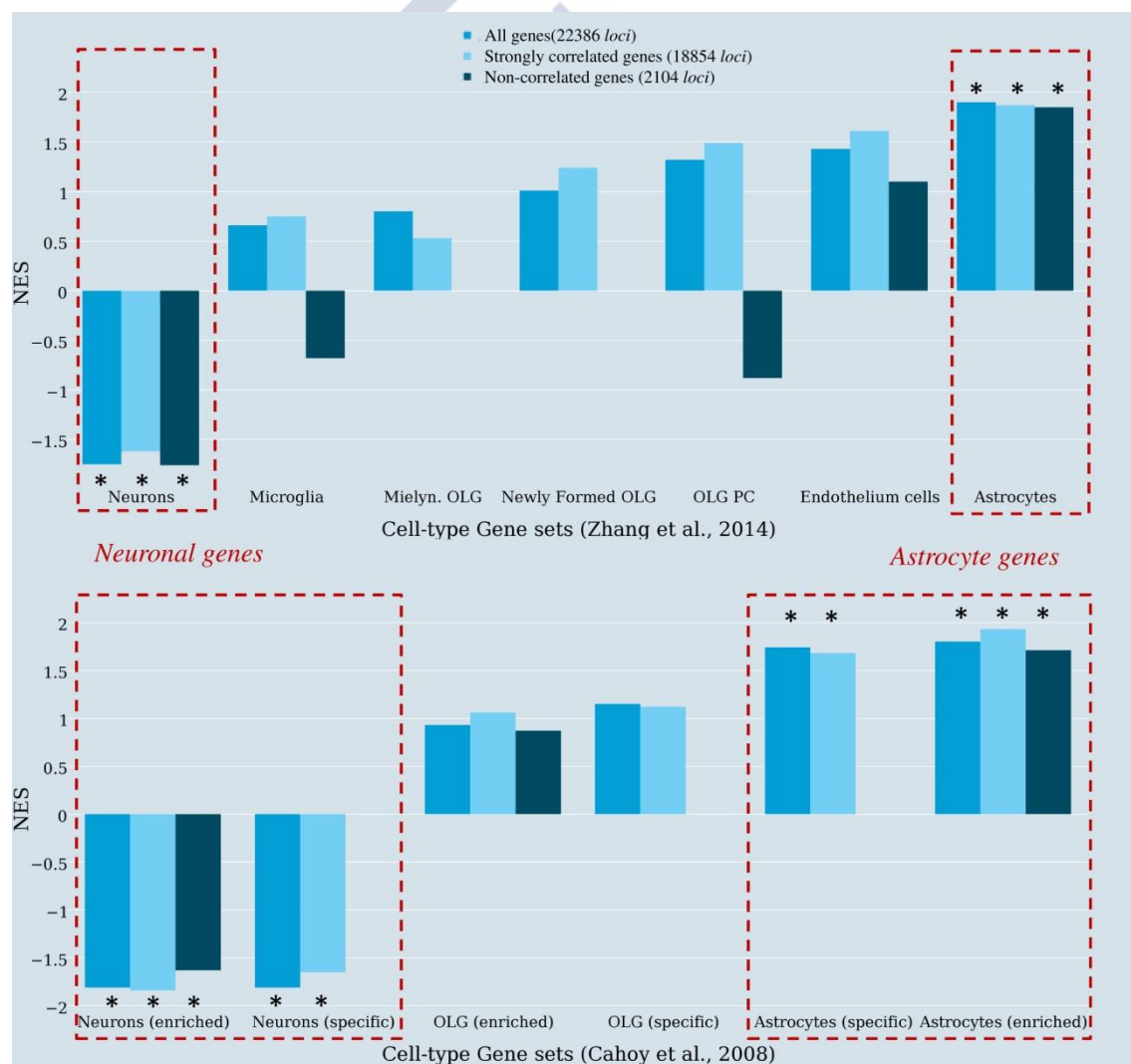
IV.2.3.2. Análisis de enriquecimiento en genes específicos de tipo celular

Los módulos sobreexpresados, M10 y C38, y subexpresados en esquizofrenia, M1 y M2, se encuentran enriquecidos en genes específicos de astrocitos y genes neuronales, respectivamente (**figura 4.14**). Para consolidar esta asociación, se estudió, mediante GSEA, el enriquecimiento de genes específicos de tipo celular cerebral concreto. Para ello se emplearon grupos de genes característicos de tipos celulares específicos de estudios previos de microarrays (Cahoy y cols., 2008) y de RNA-seq (Zhang y cols., 2014a). De forma coherente con lo reportado en el análisis de los módulos, el análisis de GSEA para tipo celular reveló una sobreexpresión significativa de aquellos genes característicos de astrocitos y una subexpresión de genes neuronales, con o sin tener en cuenta la correlación con la medicación (**figura 4.15; ANEXO VIII-IX**).

En los datos de expresión de genes no correlacionados, igualmente, se encontró una regulación al alza de los genes de astrocitos descritos tanto en el estudio de Cahoy y cols. ($N=156$ genes; $NES=1,71$; $FDR-q=0,037$) como en el de Miller y cols. ($N=45$ genes;

NES=1,85; FDR-q=0,008) y, de forma opuesta, se halló una subexpresión estadísticamente significativa de los genes característicos de neuronas descritos tanto en el estudio de Cahoy y cols. (N=133 genes; NES=-1,63; FDR-q=0,029) como en el de Miller y cols. (N=45 genes; NES=-1,76; FDR-q=0,024) (**figura 4.15**). En ningún caso se encontró enriquecimiento significativo en genes de oligodendrocitos ni microglía, pudiendo restringirse esta expresión diferencial a los tipos celulares de neuronas y de astrocitos.

Figura 4. 15. Resultados del análisis mediante GSEA del enriquecimiento de tipo celular específico en genes expresados (RPKM > 0,3). En la figura se representa el valor de enriquecimiento normalizado (NES), ordenado de menor a mayor (de izquierda a derecha), de los tipos celulares (Cahoy y cols., 2008; Zhang y cols., 2014a) en los datos de expresión de la totalidad de los genes (22386), de genes fuertemente correlacionados (18854) y de genes no correlacionados con la medicación (2104). Los tipos celulares significativamente enriquecidos (FDR-q) se encuentran marcados con un asterisco *. Mielyn. OLG = Oligodendrocito mielinizante; Newly Formed OLG = Oligodendrocito recién formado; OLG PC = Célula precursora de oligodendrocitos. Los valores de enriquecimiento y valores p, así como las gráficas de enriquecimiento de GSEA se muestran en **ANEXO VIII-XI**.



IV.2.3.3. Análisis de enriquecimiento en genes relacionados con la sinapsis

Dada la demostrada relación de la sinapsis en el desarrollo y patología de la esquizofrenia (Mirnics y cols., 2001; Owen y cols., 2005b), se estudió el enriquecimiento de los cuatro grupos de genes sinápticos (densidad postsináptica, vesículas sinápticas, densidad presináptica y zona activa presináptica) descritos en la base de datos de SynaptomeDB (Pirooznia y cols., 2012), a lo largo de los 3 grupos de genes descritos previamente (**figura 4.16**). Los cuatro conjuntos de genes sinápticos se hallaron claramente subexpresados en esquizofrenia (en todos ellos NES < -1,5), y, particularmente, los genes de la zona activa presináptica se encontraron significativamente subexpresados en esquizofrenia incluso en los datos de expresión de genes no correlacionados con la medicación (N=24 genes; NES=-1,78; FDR-q=0.003). A raíz de estos resultados, se puede inferir que, pese a que los genes del módulo C38 se encuentran enriquecidos en proteínas de la densidad postsináptica, los genes implicados en procesos sinápticos presentan una mayoritaria subexpresión en esquizofrenia, de igual modo que lo estaban los genes de módulos característicos de la expresión neuronal. Sin embargo, es precisamente en el grupo de densidad postsináptica donde no existe significación en el enriquecimiento, que se puede explicar por la presencia de un pequeño grupo de genes altamente sobreexpresados y enriquecido en genes del módulo C38 (**ANEXO IX, apartado 5.4**)

La coincidente subexpresión en los genes característicos de neuronas y aquellos implicados en procesos de sinapsis tiene sentido en tanto que éstos últimos, pese a no ser privativos del tipo celular neuronal, están enriquecidos en mayor medida en genes propios de neuronas en comparación con los de astrocitos, que es la especie celular cuyos genes se hallaron sobreexpresados en las muestras analizadas (**figuras 4.14-4.15**). 154 genes de los 678 neuronales son propios de procesos de sinapsis mientras que 74 de un total de 479 en el caso de los de astrocitos (valor P de la prueba exacta de Fisher de dos colas = 0.0021) (**tabla 4.10**).

Sin embargo, dado el mayor número de genes de densidad postsináptica en comparación con el resto, existen hasta 67 genes propios de astrocitos (más que en cualquier otra categoría), por lo que la naturaleza de este grupo de genes, comúnmente asociados con el desarrollo de la esquizofrenia (Kirov y cols., 2012; Purcell y cols., 2014), podría ser particularmente heterogénea, donde otros tipo celulares también interviniesen de forma relevante.

Resultados

Figura 4. 16. Resultados del análisis mediante GSEA del enriquecimiento de genes de sinapsis en los datos de expresión (RPKM > 0,3). En la figura se representa el valor de enriquecimiento normalizado (NES) de los 4 grupos de genes sinápticos de la base de datos de SynaptomeDB (de izda. a derecha: vesículas sinápticas, zona activa presináptica, densidad presináptica y densidad postsináptica) en los datos de la totalidad de los genes expresados (22386), genes fuertemente correlacionados (15854) y genes no correlacionados con la medicación (2104). Los valores y gráficas de enriquecimiento se muestran en ANEXO IV-VII.

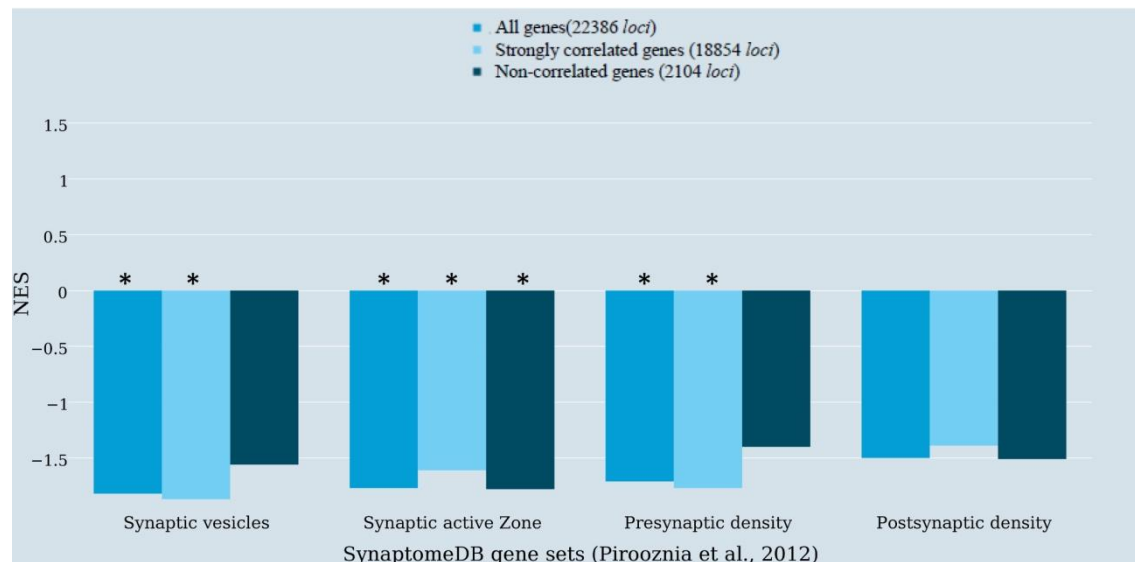


Tabla 4. 10. Enriquecimiento de genes de sinapsis en aquellos característicos de neuronas y astrocitos. Se analizó el solapamiento entre genes enriquecidos en el tipo celular (Astrocitos, N = 479; neuronas, N = 678) descritos previamente (Zhang y cols., 2014a) en aquellos de cada uno de los grupos de genes de sinapsis descritos en SynaptomeDB (Pirooznia y cols., 2012). Los valores de significación se calcularon con la comparación de proporciones de estos solapamientos con el total de genes del tipo celular mediante test exacto de Fisher de dos colas.

Tipo celular (Zhang y cols., 2014a)	Genes en SynaptomeDB				
	Densidad presináptica	Densidad postsináptica	Zona activa presináptica	Vesículas sinápticas	TODOS
	336	1755	209	107	1887
Neuronas (678)	34	142	38	21	154
%	10,10%	8,10%	18,20%	19,60%	8,20%
Astrocitos (479)	17	67	10	0	74
%	5,10%	3,80%	4,80%	0,00%	3,90%
valor P (Fisher 2- colas)	0,248	0,0025	0,0027	0,000015	0,0021

IV. 3. Genes diferentemente expresados (DE) en giro cingulado en esquizofrenia.

V.3.1. Descripción de los genes DE en esquizofrenia

Un total de 1876 *loci*, de los 22386 con expresión detectable (RPKM > 0,3), se hallaron diferentemente expresados (DE) en esquizofrenia al alcanzar significación estadística para el

análisis de expresión diferencial entre casos y controles en los *software* utilizados (FDR-p < 0,05 en EdgeR y DESeq). 1133 de los 1876 genes DE (60,4%) en esquizofrenia y los 743 restantes (39,6%) se encontraron sobreexpresados y subexpresados en las muestras cerebrales de casos respecto a controles, respectivamente. En la **tabla 4.11** se recoge el número de *loci* DE resultantes al aumentar los requisitos de significación (de valores P de 0,05 a 5×10^{-4}) a través del total de genes DE (1876 *loci*), genes fuertemente correlacionados con la medicación (1585 *loci*) y genes no correlacionados con la medicación (174 *loci*).

Tabla 4. 11. Número de *loci*, de los 22386 con expresión detectable (RPKM > 0,3), que se encuentran DE en casos y controles. Éstos datos se obtienen teniendo en cuenta la totalidad de los mismos o únicamente aquellos fuertemente correlacionados con la medicación (BY-p < 5×10^{-5}) o independientes de ella (BY-p > 0,05). Varios niveles de significación para los genes DE en ambos *software* utilizados (EdgeR y DESeq) son considerados: 0,05, 0,01, 0,005 y 0,0005.

	Todos los genes DE				Genes DE no correlacionados con medicación (BY-p > 0,05)				Genes DE fuertemente correlacionados con medicación (BY-p < 5×10^{-5})			
FDR-p	Loci DE	Sub	Sobr	% Sobr	Loci DE	Sub	Sobr	% Sobr	Loci DE	Sub	Sobr	% Sobr
0,05	1876	743	1133	60%	174	79	95	55%	1585	605	980	66%
0,01	854	327	527	61%	76	35	41	54%	729	268	461	63%
0,005	603	231	372	62%	64	27	37	58%	507	189	318	63%
0,0005	171	70	101	59%	12	4	8	67%	150	60	90	60%

Una mayoría de genes DE presenta una expresión génica correlacionada con la medicación. Excluyendo los genes DE, existe un total de 17269 y 1930 *loci* fuertemente correlacionados (BY-p < 5×10^{-5}) y *loci* no correlacionados (BY-p > 0,05) con la medicación, respectivamente, lo que representa una proporción similar de genes afectados por la medicación a lo largo de la vida, bien sea en el total de genes no DE con expresión detectable (20510 *loci*) o en el conjunto de genes DE (valor P de la prueba χ^2 de Pearson - Yates= 0,766).

Tabla 4. 12. Naturaleza de los genes correlacionados y no correlacionados con la medicación antipsicótica a lo largo de la vida. En la tabla se representa el número de *loci* que son genes codificantes o contienen al menos algún gen codificante de todos los *loci* analizados o de aquellos con expresión detectable.

Correlación con medicación	N loci	loci codificantes	loci no codificantes	% loci codificantes
DE no correlacionados	174	148	26	85,06%
DE Directamente correlacionados	815	688	127	84,42%
DE Inversamente correlacionados	770	654	116	84,94%
Todos los genes DE (exp. detectable)	1876	1585	291	84,49%

Del mismo modo, tal y como ocurría en el caso del total de genes expresados (**tabla 4.9**), la naturaleza codificante de los genes DE es homogénea a lo largo de los genes correlacionados y no correlacionados con la medicación (**tabla 4.12**). Sin embargo, sí que existe una mayor proporción de genes codificantes en aquellos *loci* DE respecto a los 20510 que no lo son (84,5% frente a 58,9%; valor P de la prueba χ^2 de Pearson ≈ 0). En cualquier caso, un total de 291 *loci* DE en este estudio son de naturaleza no codificante.

El análisis del top 15 de genes DE más significativos (**tabla 4.13**) reveló muchos previamente implicados en la etiología de la esquizofrenia (**apartado 5.4**). Del mismo modo, se analizó el top 15 de genes DE más significativos dentro de los no correlacionados con la medicación (**tabla 4.14**), hallando genes igualmente implicados en la patología de la enfermedad. En ambos casos, el total de genes descritos fueron codificantes. Cabe destacar la presencia, en tercer lugar del gen de la internexina INA, uno de los genes asociados consistentemente en los últimos estudio de GWAS en esquizofrenia (Ripke y cols., 2013a; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

Tabla 4. 13. Lista de los 15 genes con diferencias de expresión más significativas entre casos y controles. En la tabla se describe la expresión media en RPKM de casos (SCZ) y controles (CO), el P valor corregido del programa DESeq y el P valor corregido del estudio de correlación con medicación

GENCODE v19 <i>loci</i>	Exp (SCZ)	Exp (CO)	P-adj (DESeq)	P-BY (Correlación)
<i>NEUROD6</i>	58,74	93,10	$7,35 \times 10^{-8}$	$3,51 \times 10^{-56}$
<i>PVALB</i>	12,42	27,60	$8,47 \times 10^{-8}$	$1,20 \times 10^{-89}$
<i>TAC1</i>	19,89	33,43	$1,76 \times 10^{-6}$	$3,18 \times 10^{-118}$
<i>STK17B</i>	13,70	7,36	$2,97 \times 10^{-6}$	$1,55 \times 10^{-17}$
<i>PTTG1IP</i>	120,24	82,21	$3,19 \times 10^{-6}$	$5,82 \times 10^{-68}$
<i>UQCRH</i>	189,49	250,31	$4,50 \times 10^{-6}$	$1,66 \times 10^{-29}$
<i>NDUFA4</i>	246,90	336,78	$4,50 \times 10^{-6}$	$4,35 \times 10^{-81}$
<i>KCNKI</i>	101,35	143,49	$7,39 \times 10^{-6}$	$1,96 \times 10^{-136}$
<i>SMTN</i>	31,78	19,42	$7,39 \times 10^{-6}$	$2,43 \times 10^{-28}$
<i>RCAN2</i>	296,23	428,53	$7,39 \times 10^{-6}$	$6,93 \times 10^{-146}$
<i>TIMM17A</i>	69,88	104,21	$9,79 \times 10^{-6}$	$1,69 \times 10^{-89}$
<i>SEMA4B</i>	74,64	44,99	$1,35 \times 10^{-5}$	$3,90 \times 10^{-49}$
<i>RASL12</i>	20,46	11,23	$1,35 \times 10^{-5}$	$1,56 \times 10^{-83}$
<i>ALDH1L1</i>	101,38	46,20	$1,35 \times 10^{-5}$	$1,03 \times 10^{-120}$
<i>GAD2</i>	111,28	166,73	$1,62 \times 10^{-5}$	$5,67 \times 10^{-117}$

Tabla 4. 14. Lista de los 15 genes no correlacionados con medicación con diferencias de expresión más significativas entre casos y controles. En la tabla se describe la expresión media en RPKM de casos (SCZ) y controles (CO), el P valor corregido del programa DESeq y el P valor corregido del estudio de correlación.

GENCODE v19 loci	Exp (SCZ)	Exp (CO)	P-adj (DESeq)	P-BY (Correlación)
<i>NHP2L1</i>	206,81	267,60	$7,26 \times 10^{-5}$	0,08
<i>PAX6</i>	54,88	36,04	$8,98 \times 10^{-5}$	1
<i>INA</i>	246,58	346,96	$9,31 \times 10^{-5}$	0,16
<i>KIF19</i>	17,46	5,37	$1,5 \times 10^{-4}$	1
<i>CSPG4</i>	37,01	25,23	$1,35 \times 10^{-4}$	1
<i>C10orf105</i>	6,45	2,98	$1,49 \times 10^{-4}$	1
<i>ATP1B1</i>	1138,48	1528,17	$1,64 \times 10^{-4}$	1
<i>RALGAPA2</i>	63,83	47,42	$1,71 \times 10^{-4}$	1
<i>PIEZO1</i>	41,33	30,58	$1,76 \times 10^{-4}$	0,21
<i>PSAT1</i>	112,71	72,76	$1,99 \times 10^{-4}$	1
<i>TGIF1</i>	8,68	5,16	$2,51 \times 10^{-4}$	0,20
<i>CPLX2</i>	764,43	1009,55	$4,85 \times 10^{-4}$	0,54
<i>ENTPD2</i>	13,92	7,42	$5,19 \times 10^{-4}$	0,14
<i>UGP2</i>	241,96	314,95	$5,45 \times 10^{-4}$	0,41
<i>PRDX5</i>	216,91	259,71	$7,56 \times 10^{-4}$	0,08

IV.3.2. Expresión relativa de genes DE subexpresados y sobreexpresados en esquizofrenia

A fin de comprender la naturaleza de los genes sobreexpresados y subexpresados en esquizofrenia, se compararon los niveles medios de expresión de ambos conjuntos de 1133 y 743 loci, respectivamente, en el total de 57 muestras analizadas de casos y controles.

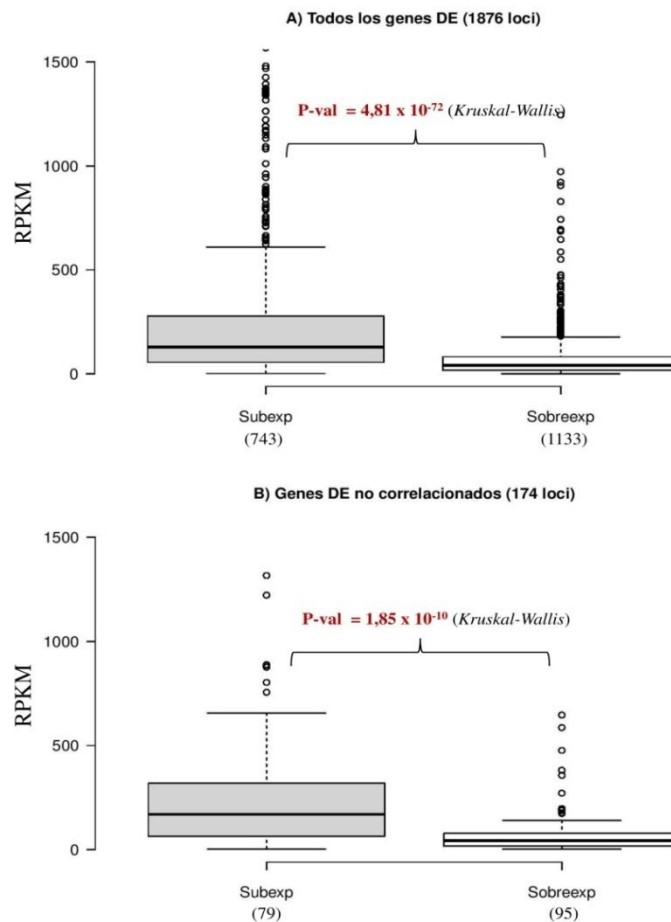
Este análisis demostró que los genes DE sobreexpresados en esquizofrenia, con independencia de la influencia de la medicación antipsicótica en las diferencias entre casos y controles, presentan unos niveles de expresión marcadamente reducidos en el presente estudio ($74,15 \pm 152,83$ en todos los genes DE sobreexpresados; $77,10 \pm 113,53$ en los genes DE sobreexpresados no correlacionados con la medicación) en comparación con los genes DE subexpresados ($250,65 \pm 377,86$ en todos los genes DE subexpresados; $274,08 \pm 322,06$ en los genes DE subexpresados no correlacionados con la medicación) (**figura 4.17**)

IV.3.3. Enriquecimiento de genes DE en tipos celulares y tejidos específicos durante el desarrollo

Con el fin de profundizar en los resultados anteriores de GSEA relativos a la subexpresión y sobreexpresión de genes de neuronas y astrocitos, respectivamente, se realizó un análisis de enriquecimiento de los genes DE en aquellos característicos de los distintos tipos

celulares cerebrales y de los distintos tejidos durante varias etapas del neurodesarrollo (**figura 4.13**). Para ello, se empleó la herramienta CSEA (del inglés, *Cell-type Specific Enrichment Analysis*) diseñada mediante análisis de datos RNA-seq de BrainSpan (Xu y cols., 2014) y de libre disposición online, con la que se estudió el enriquecimiento de las listas de genes DE en los diferentes tipos celulares y regiones cerebrales a través del desarrollo en varios niveles de especificidad (psi, del inglés *probability specificity index*) (Dougherty y cols., 2010).

Figura 4. 17. Expresión media de genes sobreexpresados y subexpresados en esquizofrenia. A) Comparación entre todos los genes DE; B) Comparación entre genes DE no correlacionados con la medicación. Entre paréntesis se indica el número de *loci* en cada caso.



Se exploró el enriquecimiento espacio-temporal de la totalidad de los genes DE (1876 *loci*) y en los genes DE no correlacionados con la medicación (174 *loci*). Además, dado que el estudio mediante GSEA describió una funcionalidad muy diferenciada de los genes regulados al alza y a la baja en esquizofrenia en el presente estudio, se realizaron las pruebas por separado en los genes DE sobreexpresados y subexpresados. Se consideraron, por un lado, todos los genes

DE (1133 *loci* sobreexpresados, 743 subexpresados) y por otro, los genes DE no correlacionados con medicación (95 *loci* sobreexpresados, 79 subexpresados).

IV.3.3.1. Análisis de enriquecimiento en tipos celulares cerebrales

Los análisis de CSEA para tipo celular confirmaron los resultados anteriores de GSEA pero en un escenario más detallado, donde los genes característicos de neuronas aparecieron subexpresados (**figura 4.20; ANEXOXXIV**) y los genes propios de células gliales, especialmente astrocitos, sobreexpresados (**figura 4.18; ANEXOXX**). Mediante este análisis, se confirma la clara separación funcional que existe entre el grupo de genes subexpresados y el de sobreexpresados. En el primer caso, las neuronas Pnoc+, Ntsr+ (Neuronas de la capa 5b) y Glt25d2 (Neuronas de la capa 6) son las que más afectadas se encuentran; mientras que la especificidad de la sobreexpresión es más pronunciada y afecta principalmente a astrocitos corticales en todos los niveles de especificidad descritos y, en general, a todas las células gliales (**ANEXO XII**)

Al considerar exclusivamente los genes no correlacionados con la medicación, aquellos sobreexpresados (95 *loci*) se encontraban significativamente enriquecidos en genes de astrocitos corticales ($\text{FDR-q} = 8,01 \times 10^{-4}$ en el umbral $\text{pSI} < 0,01$) y en menor medida, en glía de Bergman y oligodendrocitos de cerebelo ($\text{FDR-q} = 0,007$ en el umbral $\text{pSI} 0,05$) y astrocitos de cerebelo ($\text{FDR-q} = 0,035$ en el umbral $\text{pSI} 0,05$). Estos análisis de CSEA demuestran, incluso en un número reducido de genes, que los cambios de expresión observados en estos tipos celulares son una consecuencia propia de la enfermedad e independiente de la medicación administrada (**figura 4.19**).

Por otro lado, sin embargo, ningún tipo celular se encuentra enriquecido en genes DE subexpresados no correlacionados con la medicación (**figura 4.21**). Es necesario señalar, no obstante, que tan solo 79 *loci* DE se hallaron subexpresados, lo cual reduce la potencia estadística considerablemente respecto al análisis en GSEA, que sí reveló una subexpresión independiente de la medicación.

Figura 4. 18. Análisis de CSEA para el enriquecimiento en tipo celular de los genes DE sobreexpresados en esquizofrenia. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 852 genes presentes en la base de datos de BrainSpan pertenecientes a los 1133 *loci* sobreexpresados. En el ANEXOXX se encuentran todos los valores de significación para cada uno de los casos analizados y en todos los tipos celulares considerados.

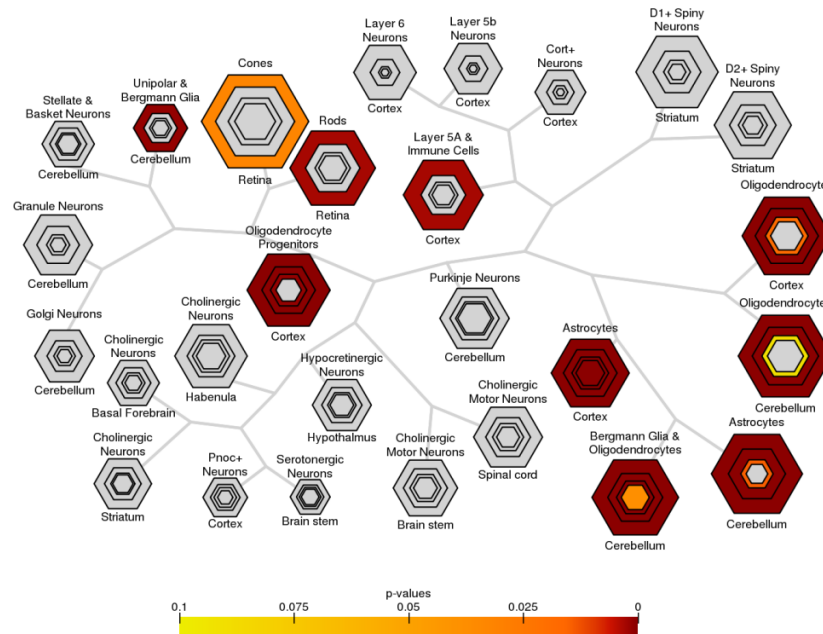


Figura 4. 19. Análisis de CSEA para el enriquecimiento en tipo celular de los genes DE sobreexpresados en esquizofrenia y no correlacionados con la medicación. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 71 genes presentes en la base de datos de BrainSpan pertenecientes a los 95 *loci* sobreexpresados. En el ANEXOXX se encuentran todos los valores de significación para cada uno de los casos analizados y en todos los tipos celulares considerados.

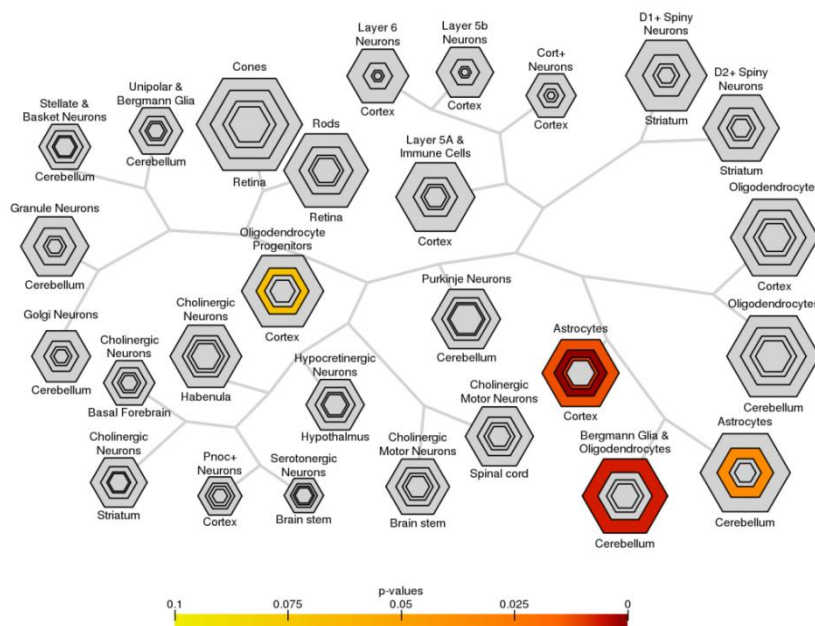


Figura 4. 20. Análisis de CSEA para el enriquecimiento en tipo celular de los genes DE subexpresados en esquizofrenia. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 556 genes presentes en la base de datos de BrainSpan pertenecientes a los 743 *loci* subexpresados. En el **ANEXOXXIV** se encuentran todos los valores de significación para cada uno de los casos analizados y en todos los tipos celulares considerados.

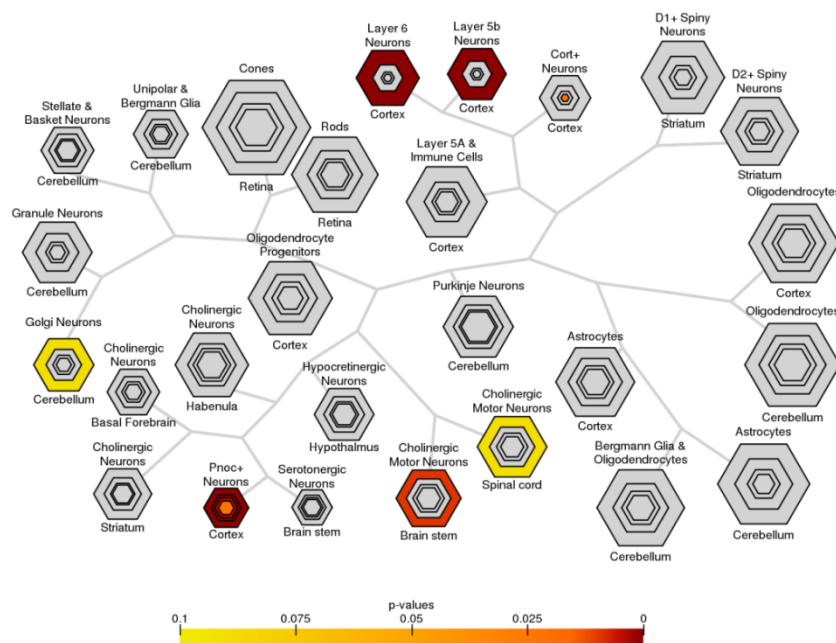
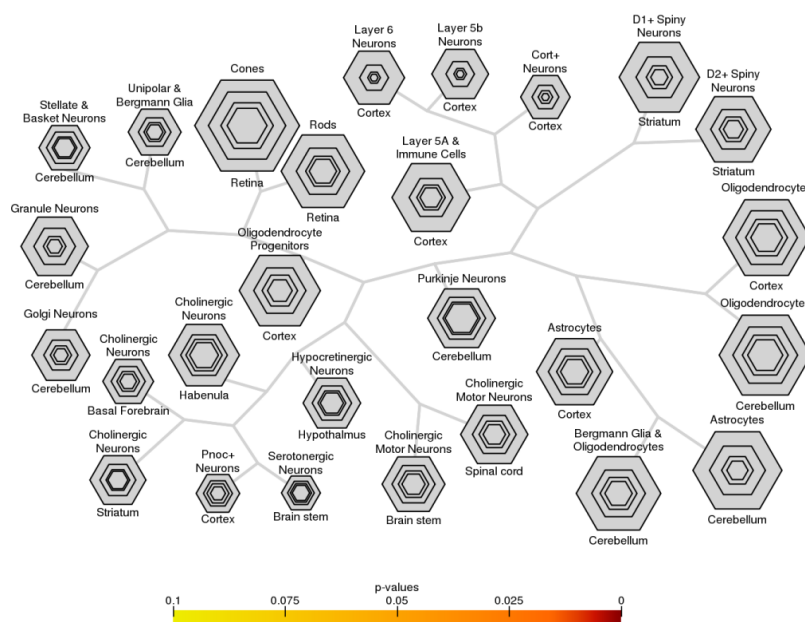


Figura 4. 21. Análisis de CSEA para el enriquecimiento en tipo celular de los genes DE subexpresados en esquizofrenia y no correlacionados con la medicación. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 61 genes presentes en la base de datos de BrainSpan pertenecientes a los 79 *loci* subexpresados. En el **ANEXOXXIV** se encuentran todos los valores de significación para cada uno de los casos analizados y en todos los tipos celulares considerados.



IV.3.3.2. Análisis de enriquecimiento en áreas cerebrales durante el desarrollo

Respecto al estudio de áreas cerebrales durante el desarrollo, el análisis de CSEA reveló que la influencia de la medicación es muy relevante en la alteración de los niveles de expresión de genes característicos de las diferentes etapas del neurodesarrollo (**figuras 4.22 -4.24**), haciendo muy complicada su interpretación.

En líneas generales, se observa una tendencia a la sobreexpresión de los genes específicos de las áreas cerebrales analizadas en las etapas tempranas del desarrollo, particularmente importante en el caso del cerebelo (**figura 4.22; ANEXO XXXII**), mientras que se observa una tendencia a la subexpresión de los genes específicos de etapas tardías del desarrollo, principalmente en córtex y tálamo (**figura 4.24; ANEXO XXXVI**).

Sin embargo, también se demostró el enriquecimiento independiente de la medicación de los genes DE en regiones concretas durante el desarrollo, (**figuras 4.23 y 4.25**). Los genes subexpresados en esquizofrenia se hallaron particularmente enriquecidos en aquellos característicos del tálamo durante la infancia media ($\text{FDR-q} = 0,008$ en el umbral $\text{pSI } 0,05$) y tardía ($\text{FDR-q} = 0,008$ en el umbral $\text{pSI } 0,05$), del hipocampo durante la niñez media ($\text{FDR-q} = 0,008$ en el umbral $\text{pSI } 0,05$) y de la corteza en edad adulta ($\text{FDR-q} = 0,008$ en el umbral $\text{pSI } 0,05$). Por contrapartida, los genes sobreexpresados en esquizofrenia se encontraron enriquecidos en aquellos específicos de hipocampo durante la etapa fetal temprana ($\text{FDR-q} = 0,018$ en el umbral $\text{pSI } 0,05$) e infancia temprana ($\text{FDR-q} = 7,56 \times 10^{-7}$ en el umbral $\text{pSI } 0,05$). La mayor significación se encontró en genes expresados específicamente en hipocampo durante la infancia temprana, que se vieron altamente enriquecidos incluso cuando se consideraron todos los genes DE no correlacionados (**ANEXOXXXVIII**; $\text{FDR-q} = 1,53 \times 10^{-4}$ en el umbral $\text{pSI } 0,05$).

Figura 4. 22. Análisis de CSEA para el enriquecimiento en región cerebral durante el desarrollo de los genes DE sobreexpresados en esquizofrenia. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 959 genes presentes en la base de datos de BrainSpan pertenecientes a los 1133 *loci* sobreexpresados. En el **ANEXOXXXII** se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.

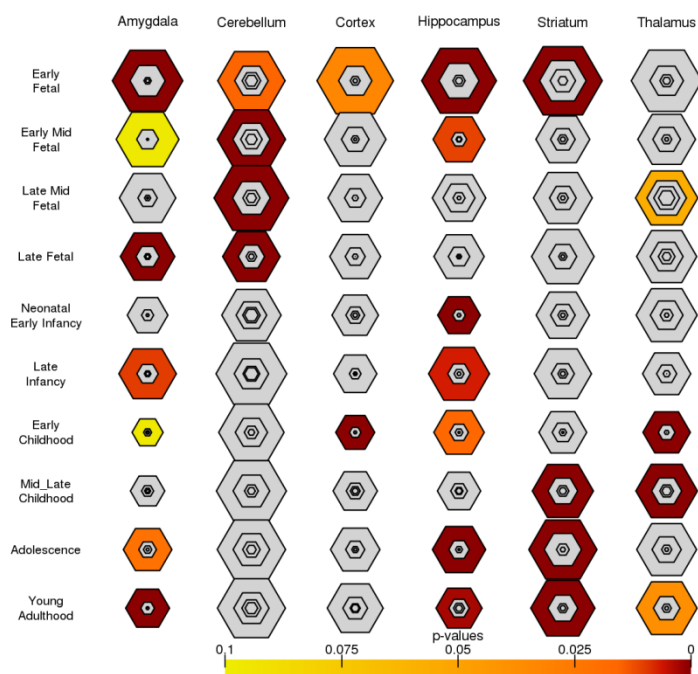


Figura 4. 23. Análisis de CSEA para el enriquecimiento en región cerebral durante el desarrollo de los genes DE sobreexpresados en esquizofrenia y no correlacionados con la medicación. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 86 genes presentes en la base de datos de BrainSpan pertenecientes a los 95 *loci* sobreexpresados. En el **ANEXOXXXII** se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.

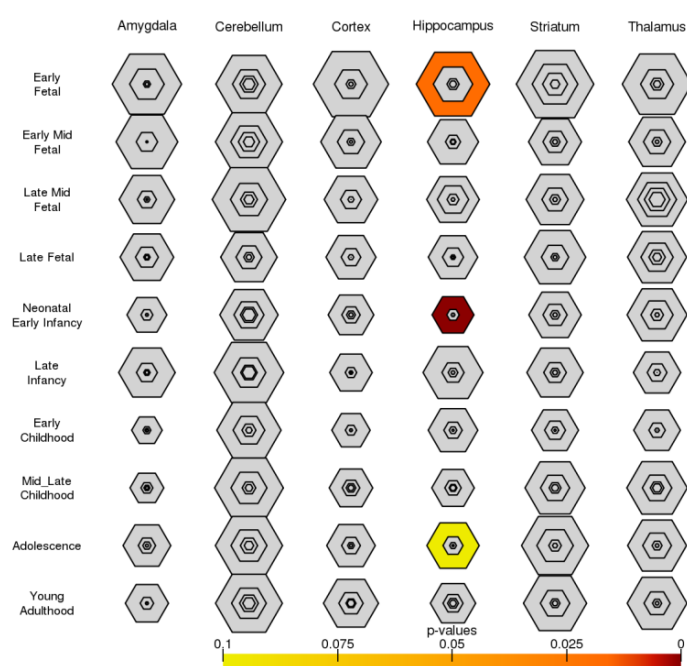


Figura 4. 24. Análisis de CSEA para el enriquecimiento en región cerebral durante el desarrollo de los genes DE subexpresados en esquizofrenia. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 625 genes presentes en la base de datos de BrainSpan pertenecientes a los 743 *loci* subexpresados. En el ANEXOXXXVI se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.

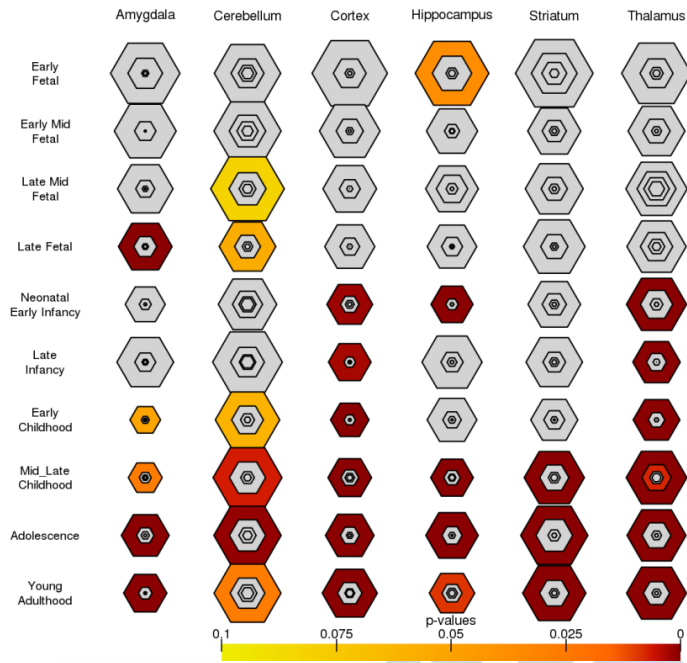
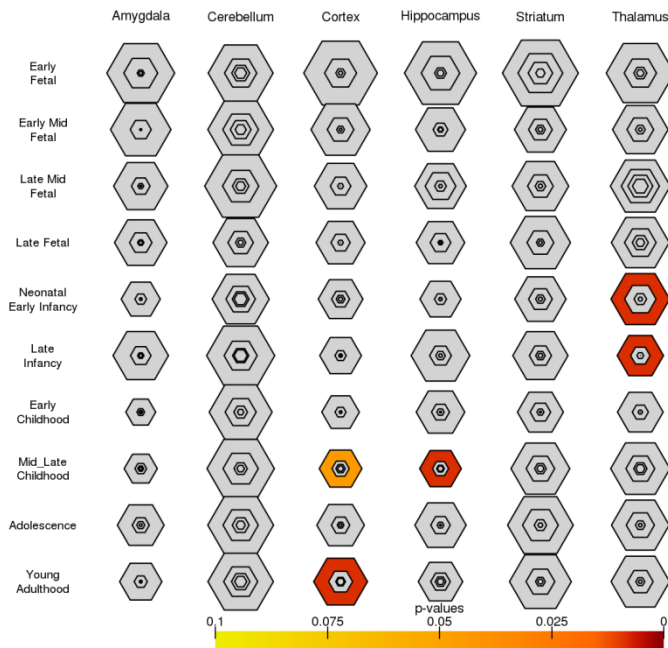


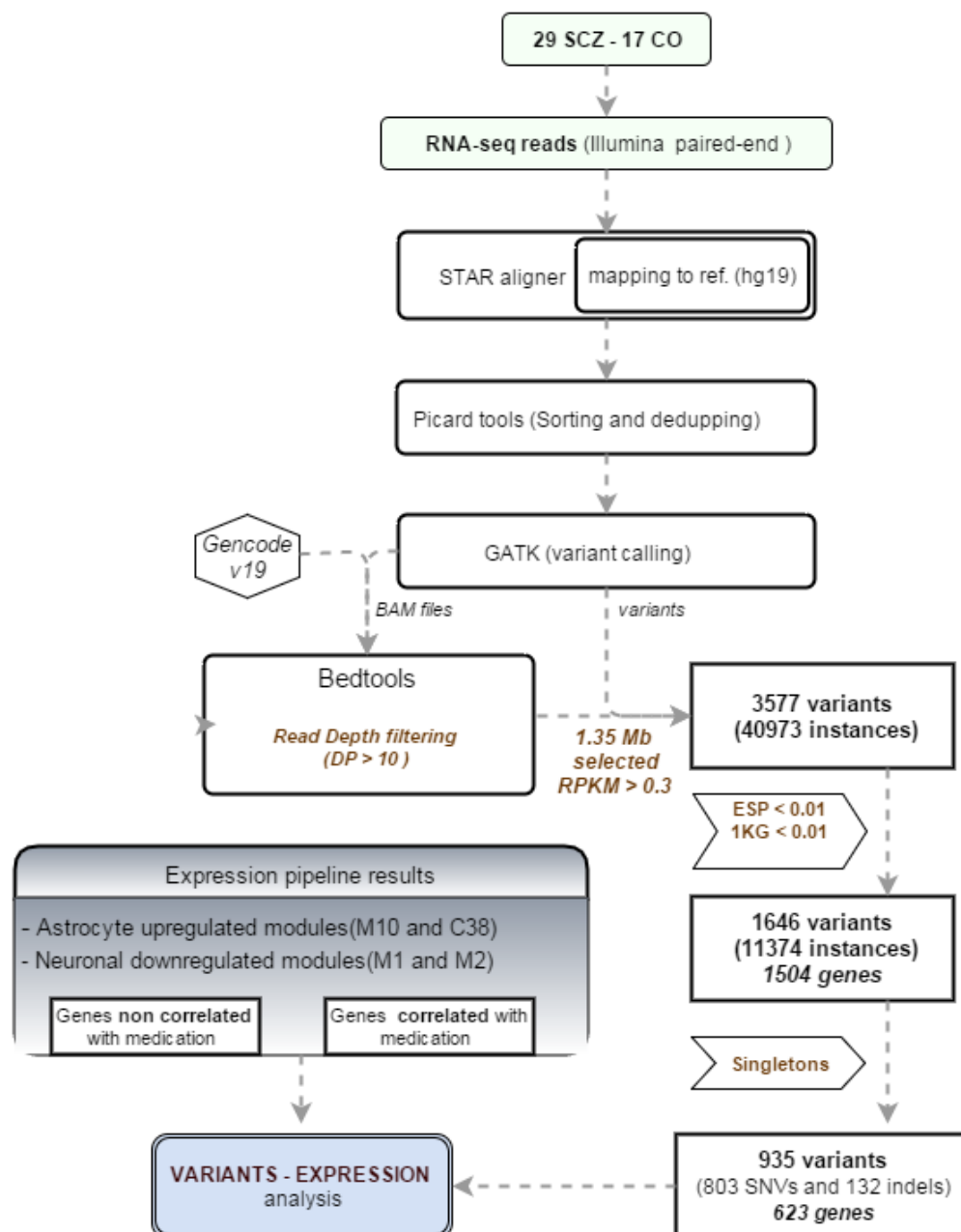
Figura 4. 25. Análisis de CSEA para el enriquecimiento en región cerebral durante el desarrollo de los genes DE subexpresados en esquizofrenia y no correlacionados con la medicación. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 70 genes presentes en la base de datos de BrainSpan pertenecientes a los 79 *loci* subexpresados. En el ANEXOXXXVI se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.



IV. 4. Análisis de variantes de genes altamente expresados en las muestras de ARN de giro cingulado.

IV. 4. 1. Análisis descriptivo de las variantes detectadas y de los genes mutados. Los genes mutados se encuentran subexpresados en esquizofrenia.

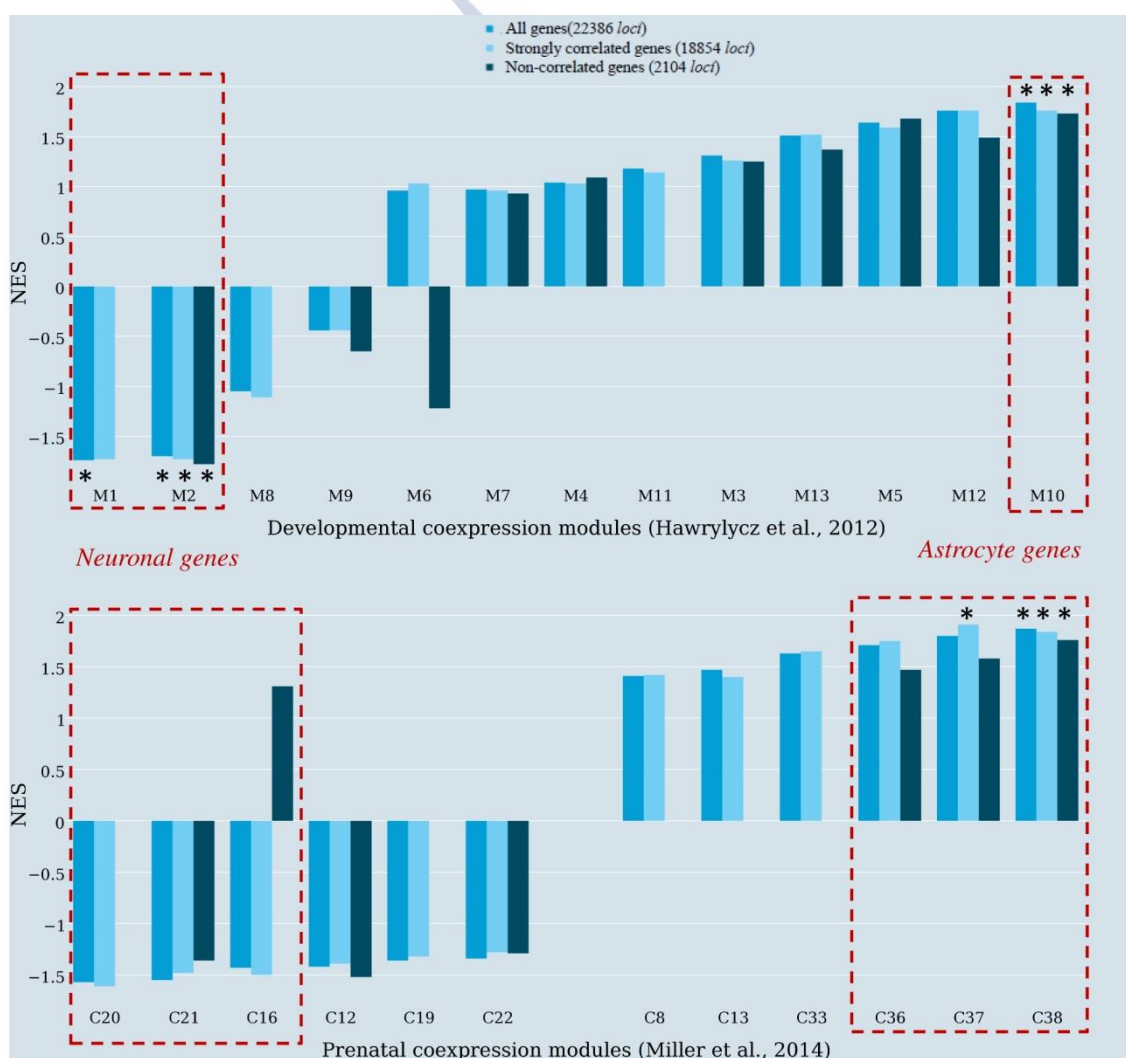
Figura 4. 26. Procedimiento de análisis de variantes raras de genes altamente expresados en 29 muestras de ARN de esquizofrenia y 17 de controles.



Resultados

Se exploró la variación genética rara (*singletons* no reportados o con frecuencia alélica $< 0,01$ en las bases de datos de ESP y de los 1000 genomas) a través de genes y posiciones genómicas expresadas de forma consistente (posiciones con DP > 10 en genes con niveles de expresión RPKM $> 0,3$) en todas las muestras de giro cingulado (**figura 4.26**).

Figura 4. 27. Análisis en GSEA del enriquecimiento de módulos de coexpresión en genes expresados (RPKM $> 0,3$) para 29 muestras de casos y 17 de controles sanos. En la figura se representa el valor de enriquecimiento normalizado (NES), ordenado de menor a mayor (de izquierda a derecha), de los módulos de coexpresión (Hawrylycz y cols., 2012; Miller y cols., 2014) en los datos de expresión de la totalidad de los genes (22386), de genes fuertemente correlacionados (15854) y de genes no correlacionados con la medicación (2104). Los módulos estadísticamente significativos tras correcciones múltiples (FDR-q) están marcados con un asterisco *. Los valores de enriquecimiento y valores p en cada caso se encuentran descritos en ANEXOXXXVII -XXXVIII. Las gráficas de enriquecimiento de GSEA se muestran en ANEXO XXXIX.



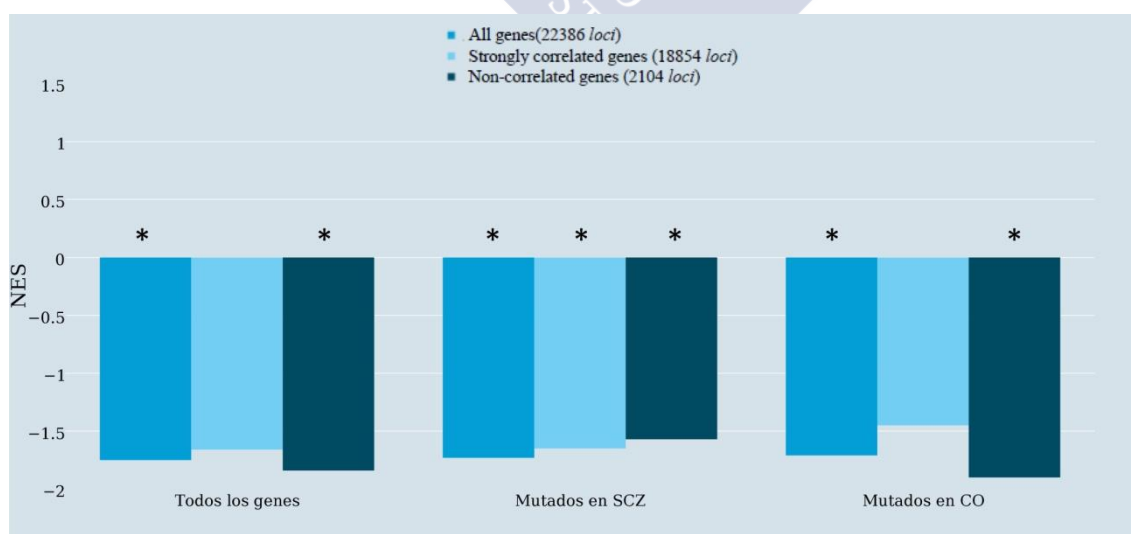
Para este análisis, 2 muestras de casos y 9 controles fueron eliminadas debido a su baja cobertura en comparación con el resto de muestras (**métodos**), por lo que los análisis de

expresión se repitieron en las 29 muestras de casos y 17 de controles utilizadas en el análisis de variantes, a fin de descartar que los resultados en el estudio de expresión fuesen producto de la presencia de estas muestras con coberturas marcadamente inferiores. La **figura 4.27** demuestra que los enriquecimientos en módulos de expresión no solo mantienen la misma tendencia, sino que guardan significación tal y como hacían en la totalidad de las muestras.

El análisis de variantes siguiendo el protocolo diseñado para tal efecto (**métodos, figura 4.26**) concluyó en la detección de 803 SNV y 132 indels, raros y *singletons*, en los transcriptomas de 29 casos y 17 controles. El filtro posicional seleccionó 1,35 Mb con DP > 10 en genes RPKM > 0,3, lo que constituye un 12% del total del transcriptoma cerebral (descrito por el *Allen Brain Atlas*, descargado desde el navegador de UCSC). En esos 1,35 Mb, el total de 935 variantes detectadas afectaron a un total de 623 genes (genes mutados).

Mediante GSEA, se estudió el perfil de expresión de todos los genes mutados (623), y de los genes únicamente mutados en casos (331) o en controles (185) a través de los datos de expresión previamente descritos (**figura 4.28**): todos los genes con expresión detectable (N = 22386), genes fuertemente correlacionados con la medicación (N = 18854) y genes no correlacionados con la medicación (N = 2104).

Figura 4. 28. Análisis en GSEA del enriquecimiento de genes mutados en genes expresados (RPKM > 0,3) para 29 muestras de casos y 17 de controles sanos. En la figura se representa el valor de enriquecimiento normalizado (NES) del total de genes mutados (623), de los mutados exclusivamente en casos (331) y de los mutados exclusivamente en controles (185) en los datos de expresión de la totalidad de los genes (22386), de genes fuertemente correlacionados (18854) y de genes no correlacionados con la medicación (2104). Los módulos estadísticamente significativos tras correcciones múltiples (FDR-q) están marcados con un asterisco *. Los valores de enriquecimiento y valores p en cada caso se encuentran descritos en **ANEXOXXXX**.



Los genes subexpresados en esquizofrenia se hallaron enriquecidos en genes portadores de variantes, tanto en el conjunto de genes como incluso cuando se utiliza el conjunto de datos no correlacionados con la medicación, para todos los genes que portan mutaciones (N = 63 genes; NES = -1,84; FDR q-val = 0,003), aquellos mutados solamente en los casos (N = 27 genes; NES = -1,57; FDR-q = 0,039) o los genes mutados únicamente en controles (N = 25 genes; NES = -1,9; FDR-q = 0,003). Curiosamente, sólo los genes únicamente mutados en casos están significativamente subexpresados al considerar el conjunto de genes fuertemente correlacionados con la medicación (N = 154 genes; NES = -1.65; FDR-q = 0,049), así como menos enriquecidos en aquellos no correlacionados (N = 27 genes; NES = -1.57; FDR-q = 0,039). Estas observaciones podrían sugerir una mayor influencia de la medicación en la expresión de los genes mutados en esquizofrenia. No obstante, en general, estas observaciones preliminares describen un escenario en el que los genes consistentemente expresados que portan mutaciones raras están consistentemente regulados a la baja en las muestras de esquizofrenia.

IV. 4. 2. Estudio de la variación genética en módulos subexpresados en esquizofrenia. Disrupción de genes implicados en el desarrollo cortical en casos de esquizofrenia.

A fin de estudiar la relación de la variación genética detectada con los datos de expresión en cerebro, se realizó un análisis de la primera en el contexto de los módulos subexpresados y sobreexpresados anteriormente estudiados cuyos resultados fuesen significativos.

En un primer lugar, se estudió la presencia de variantes raras en genes mutados pertenecientes a módulos subexpresados en esquizofrenia estableciendo las categorías siguientes: 3 conjuntos de genes expresados pertenecientes a los módulos M1 y M2 (3728 *loci* expresados, 3028 *loci* fuertemente correlacionados con la medicación y 397 *loci* no correlacionados con la medicación) y 3 conjuntos de genes DE en esquizofrenia pertenecientes a los módulos M1 y M2 (459 *loci* DE, 373 *loci* fuertemente correlacionados con la medicación y 48 *loci* no correlacionados con la medicación). Se realizó un análisis del enriquecimiento de variantes genéticas y de genes mutados pertenecientes a módulos subexpresados en esquizofrenia en el total de genes mutados con expresión detectable (RPKM > 0,3)

En coherencia con la anterior observación de que los genes DE subexpresados en esquizofrenia se encuentran altamente expresados en giro cingulado (**figura 4.17**) y de que los genes mutados se encuentran generalmente subexpresados en esquizofrenia (**figura 4.28**), el análisis de genes mutados de módulos subexpresados M1 y M2 demostró un fuerte enriquecimiento de los mismos en cada una de las categorías del total de 623 genes mutados y

expresados de forma consistente (variantes en posiciones con DP > 10 en todas las muestras de genes con RPKM > 0,3), así como en cada una de las categorías analizadas (**tabla 4.15-A**).

No obstante, no se encontró acumulación de variantes en casos frente a controles en cualquiera de los sets de genes estudiados (**tabla 4.15-A**).

Tabla 4. 15. Análisis del enriquecimiento de variantes y de genes mutados pertenecientes a módulos subexpresados (A) y sobreexpresados (B) en esquizofrenia en el total de genes mutados con expresión detectable (RPKM > 0,3). En el caso del análisis de genes mutados se realizaron pruebas χ^2 de Pearson para el enriquecimiento de los genes mutados en cada categoría de subexpresados o sobreexpresados respecto al total de genes mutados con RPKM > 0,3. Para el análisis de variantes, se comparó la acumulación de variantes en casos (SCZ) y controles (CO) de cada categoría de genes mutados subexpresados o sobreexpresados respecto a las variantes en las mismas categorías de genes mutados con RPKM > 0,3 utilizando un test exacto de Fisher de dos colas. Los valores P de ambas pruebas se corrigieron por múltiples test mediante la estimación del FDR de Benjamini-Hochberg en los 10 test implementados en cada caso.

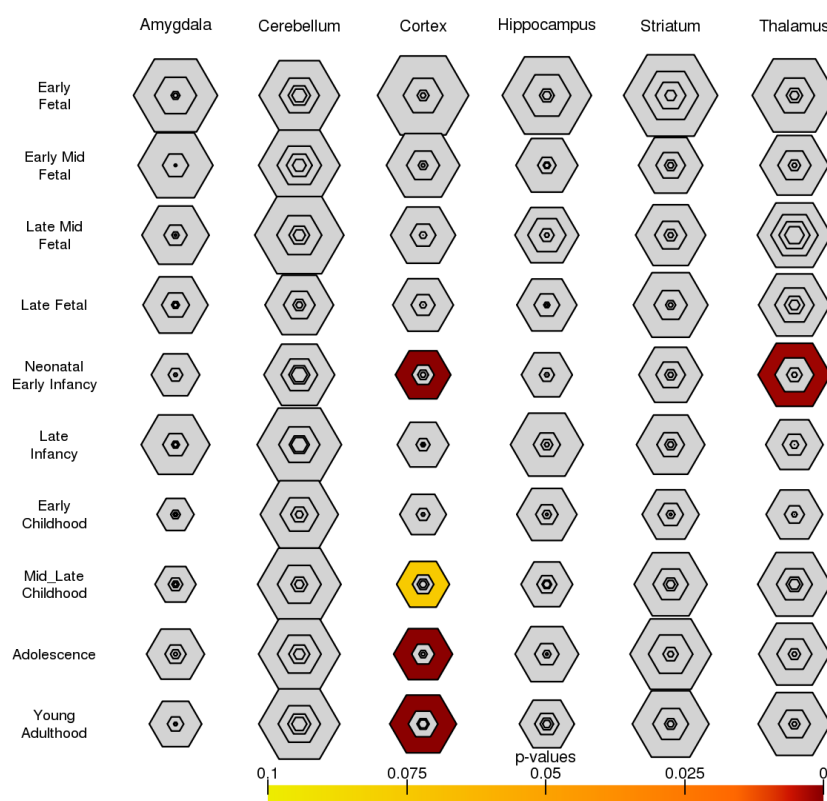
Genes RPKM > 0,3	N genes mutados (DP>10)	Enriquecimiento en n° de genes (FDR-q)	Variantes SCZ (29)	Variantes CO (17)	SCZ/CO	Acumulación de variantes (FDR-q)
Todos los genes (22386)	623		586	349	0,98	
No correlacionados (2104)	63		54	43	0,74	
Correlacionados (18854)	518		494	283	1,02	
Todos los genes DE (1876)	150		141	90	0,92	
No correlacionados DE (174)	18		15	10	0,88	
Correlacionados DE (1585)	119		109	76	0,84	
A) Genes subexpresados (Módulos M1 y M2)						
Todos los genes (3728)	345	< 10⁻¹⁵ (< 10⁻¹⁵)	318	201	0,93	0,612 (0,921)
No correlacionados (397)	36	3,2 x 10⁻¹⁵ (6,4 x 10⁻¹⁵)	27	23	0,69	0,863 (0,983)
Correlacionados (3028)	282	< 10⁻¹⁵ (< 10⁻¹⁵)	252	162	0,91	0,379 (0,758)
Todos los genes DE (459)	99	< 10⁻¹⁵ (< 10⁻¹⁵)	95	62	0,90	0,885 (0,983)
No correlacionados DE (48)	14	4,9 x 10⁻⁷ (8,1 x 10⁻⁷)	11	8	0,81	1 (1)
Correlacionados DE (373)	75	< 10⁻¹⁵ (< 10⁻¹⁵)	68	50	0,80	0,645 (0,921)
B) Genes sobreexpresados (Módulos M10 y C38)						
Todos los genes (591)	11	0,166 (0,184)	23	3	4,49	0,0063 (0,032)
No correlacionados (57)	-	-	-	-	-	-
Correlacionados (502)	11	0,435 (0,435)	23	3	4,49	0,0064 (0,032)
Todos los genes DE (185)	6	0,012 (0,017)	11	3	2,15	0,258 (0,645)
No correlacionados DE (17)	-	-	-	-	-	-
Correlacionados DE (160)	6	0,057 (0,071)	11	3	2,15	0,161 (0,537)

En segundo lugar, se propuso analizar a fondo la naturaleza de estas mutaciones en casos y controles en relación con los perfiles de expresión espaciotemporales. Para ello, y dada la naturaleza de la esquizofrenia como trastorno del neurodesarrollo, se probó el

Resultados

enriquecimiento espacio-temporal de los genes mutados en casos y en controles, exclusivamente, mediante CSEA, observándose que los genes mutados pertenecientes a los módulos subexpresados M1 y M2 (345 *loci*) se encontraban enriquecidos en genes de neuronas de proyección cortico-talámica *Ntsr+* de la corteza cerebral, bien considerando sólo genes mutados en casos (N = 177 *loci*, FDR-q = 0,042 en el umbral pSI 0,05) o en controles (N = 107 *loci*; FDRq-val = 0,013 en el umbral pSI 0,05) (ANEXO XXXXI-XXXXII). Esto refuerza la observación previa en este trabajo por la que genes mutados y expresados ubicuamente se hallan enriquecidos en neuronas, pero implicando específicamente al tipo *Ntsr+*, que se caracterizan por la presencia del receptor 1 del neuromodulador Neurotensina (NT), cuya densidad ya se vio reducida, anteriormente, en pacientes con esquizofrenia (Hamid y cols., 2002).

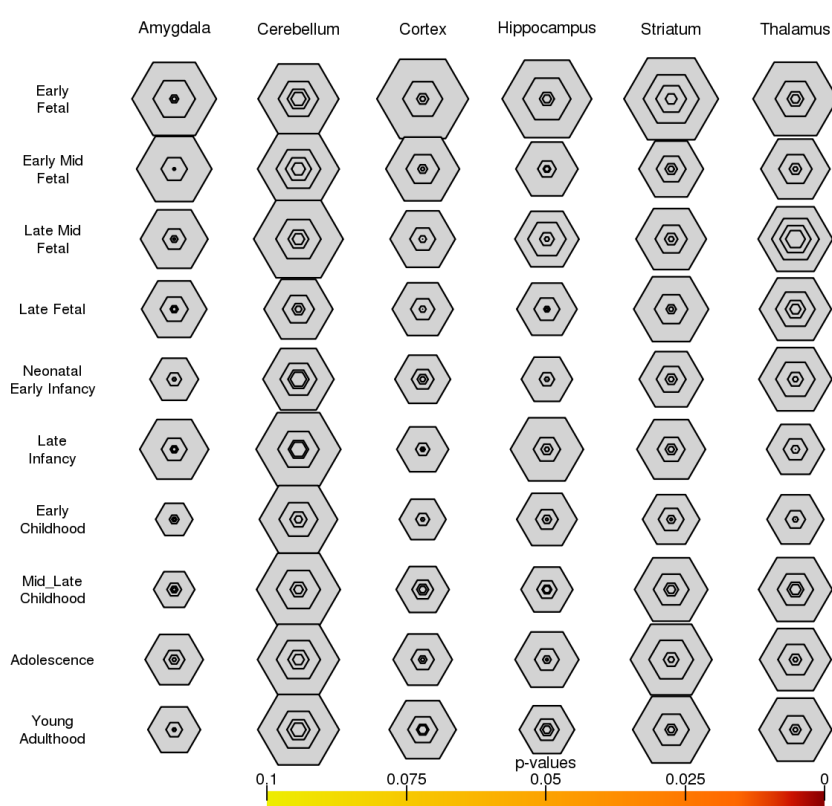
Figura 4. 29. Análisis de CSEA para el enriquecimiento en región cerebral durante el desarrollo de los genes mutados exclusivamente en casos y pertenecientes a módulos subexpresados en esquizofrenia. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 191 genes presentes en la base de datos de BrainSpan pertenecientes a los 177 *loci* mutados en casos de los módulos subexpresados en esquizofrenia M1 y M2. En el ANEXO XXXXIII se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.



Por otra parte, al probar el enriquecimiento de genes mutados a través de regiones del cerebro durante el desarrollo, se revelaron resultados diferentes en casos y en controles. Solo las

variantes presentes en genes únicamente mutados en individuos afectados se vieron acumuladas en genes corticales a través de diferentes etapas de desarrollo (**figura 4.29**): en infancia temprana ($\text{FDR-q} = 2,58 \times 10^{-4}$ en el umbral pSI 0,05), en niñez tardía ($\text{FDR-q} = 0,031$ en el umbral pSI 0,05), adolescencia ($\text{FDR-q} = 2,58 \times 10^{-4}$ en el umbral pSI 0,05) y, sobretodo, en edad adulta ($\text{FDR-q} = 4,3 \times 10^{-7}$ en el umbral pSI 0,05).

Figura 4. 30. Análisis de CSEA para el enriquecimiento en región cerebral durante el desarrollo de los genes mutados exclusivamente en controles y pertenecientes a módulos subexpresados en esquizofrenia. En la figura se muestran los valores de enriquecimiento en los umbrales de especificidad (pSI) para los 121 genes presentes en la base de datos de BrainSpan pertenecientes a los 107 *loci* mutados en casos de los módulos subexpresados en esquizofrenia M1 y M2. En el **ANEXO XXXIV** se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.



Además, la región del tálamo en infancia temprana también se halló enriquecida en genes mutados únicamente en esquizofrenia ($\text{FDR-q} = 0,001$ en el umbral pSI 0,05), lo cual es coherente con la proyección cortico-talámica de las neuronas *Ntsr+*. En contraposición, no se encontró ninguna asociación significativa al utilizar los genes mutados sólo en controles (**figura 4.30**). Igualmente, este enriquecimiento en las mismas regiones cerebrales sigue presente cuando se considera la totalidad de los genes mutados sólo en SCZ ($N = 331$ *loci*) o CO ($N = 185$ *loci*), en lugar de sólo aquellos que pertenecen a los módulos subexpresados

(ANEXOXXXV - XXXXVI), algo que concuerda con la sobrerrepresentación de los genes mutados en los módulos de coexpresión subexpresados en esquizofrenia (**tabla 4.15-A**).

IV. 4. 3. Estudio de la variación genética en módulos sobreexpresados en esquizofrenia. Acumulación de variantes raras en casos respecto a controles.

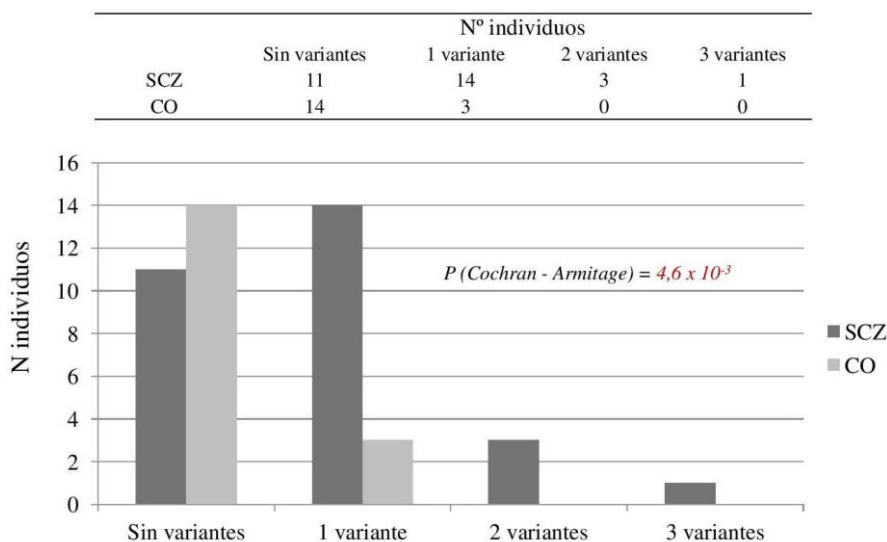
De igual modo a lo realizado en módulos subexpresados, también se estudió la presencia de variantes raras en genes mutados pertenecientes a módulos sobreexpresados en esquizofrenia estableciendo las mismas 6 categorías, aunque en este caso fueron 4 finalmente debido a la ausencia de genes de estos módulos que no se hallasen correlacionados con la medicación (**tabla 4.15-B**): 2 conjuntos de genes expresados pertenecientes a los módulos M10 y C38 (11 *loci* expresados y los mismos fuertemente correlacionados con la medicación) y 2 conjuntos de genes DE en esquizofrenia pertenecientes a los módulos M10 y C38 (6 *loci* DE y los mismos fuertemente correlacionados con la medicación).

En contraposición a lo encontrado para los módulos subexpresados en el presente trabajo, los genes de los módulos M10 y C38 presentaron un menor número de variantes de lo esperado por azar en el caso de los genes DE ($FDR-q = 0,0414$; **tabla 4.15-B**), sugiriendo que una gran parte de las posiciones genómicas de estos genes se eliminasen por los procesos de filtrado ($DP > 10$), al estar éstos menos expresados que aquellos de módulos subexpresados en esquizofrenia. Otra posibilidad sería la mayor intolerancia de estos genes sobreexpresados relativos a astrocitos a sufrir mutaciones en su secuencia. Para probar esta segunda hipótesis, se compararon los valores de intolerancia a la variación genética en la secuencia (*RVIS score*) de los genes pertenecientes a los módulos subexpresados (M1 y M2) y sobreexpresados en esquizofrenia (C38 y M10), anteriormente publicados (Petrovski y cols., 2013; Petrovski y cols., 2015). Esta comparación reveló, contrariamente a la hipótesis formulada, que los genes de los módulos subexpresados en esquizofrenia son mucho más intolerantes a la variación genética que los de los módulos sobreexpresados, tanto para variación codificante (P de Kruskal-wallis = $4,6 \times 10^{-94}$) como para variación en regiones UTR e intrónicas (P de Kruskal-wallis = $5,61 \times 10^{-83}$). Por ello, la acumulación de variantes significativamente menor en genes de módulos sobreexpresados sugiere un sesgo metodológico por el cual éstos, menos expresados en las muestras de giro cingulado, presentan un menor porcentaje de su secuencia que pasase el filtro posicional implementado ($DP > 10$ en todas las muestras).

Por otra parte, los genes pertenecientes a estos módulos sobreexpresados en esquizofrenia sí presentaron una acumulación significativa de variantes en casos en comparación con los controles (**tabla 4.15-B**), aunque ésta no alcanzó significación al incluir

exclusivamente a los genes DE. En los 11 genes mutados que pertenecen a los módulos sobreexpresados se detectaron 23 variantes en casos y 3 en controles, existiendo una tendencia significativa en el análisis de número de variantes por individuo (**figura 4.31**; valor P del test de Cochran-Armitage = $4,6 \times 10^{-3}$).

Figura 4. 31. Resultado del test de tendencia para la acumulación de variantes en casos y controles. En la tabla y el histograma inferior se describe el número de mutaciones por individuo, desde la ausencia de las mismas hasta 3 mutaciones en un individuo esquizofrénico.



Con estos resultados, se demuestra un vínculo entre genes de astrocitos regulados positivamente en esquizofrenia y la acumulación de variantes raras en casos respecto a controles. En el caso del módulo C38 (Miller y cols., 2014), se detectaron 8 variantes en casos pero ninguna en controles (**tabla 4. 16**), a lo largo de tres genes (5 mutaciones en *PCBD2*, una en *BCL2L14* y dos en *CAMTA1*).

Por otro lado, 7 genes del módulo sobreexpresado M10 (*APOE*, *ATPIA2*, *GFAP*, *LRIG1*, *NTRK2*, *PMP2* y *SLC1A2*) albergan mutaciones en 15 casos y 3 genes (*GLUD1*, *ATPIA2* y *SLC1A2*) en 3 controles (**tabla 4.17**).

Resultados

Tabla 4. 16. Variantes pertenecientes al módulo C38 sobreexpresado en esquizofrenia en 29 casos y 17 controles sanos. Todas las variantes se encuentran fuera de regiones codificantes, siendo UTR en su totalidad.

CHR	Posición inicio (hg19)	Posición final (hg19)	REF	ALT	SCZ	CO	ID	loci Gencode v.19	genes DE	dbSNP 142
chr1	7828174	7828178	AAAAA	-	1	0	A78	CAMTA1	NO	-
chr1	7828279	7828280	AA	-	1	0	A75	CAMTA1	NO	-
chr5	134259899	134259899	T	C	1	0	A100	CATSPER3, PCBD2	NO	-
chr5	134262722	134262722	T	C	1	0	A10	CATSPER3, PCBD2	NO	-
chr5	134262754	134262754	A	G	1	0	A17	CATSPER3, PCBD2	NO	rs562390790
chr5	134263461	134263461	A	G	1	0	A3	CATSPER3, PCBD2	NO	-
chr5	134263477	134263477	A	G	1	0	A12	CATSPER3, PCBD2	NO	-
chr12	12264563	12264563	G	A	1	0	A7	BCL2L14	NO	-

Tabla 4. 17. Variantes pertenecientes al módulo M10 sobreexpresado en esquizofrenia en 29 casos y 17 controles sanos. En cursiva se señalan las variantes exónicas, encontrándose el resto en regiones UTR, fuera de regiones codificantes.

CHR	Posición inicio (hg19)	Posición final (hg19)	REF	ALT	SCZ	CO	ID	loci Gencode v.19	genes DE	dbSNP 142
chr1	160111232	160111232	A	G	1	0	A81	ATP1A2	SI	-
chr1	160112341	160112341	A	G	1	0	A45	ATP1A2	SI	-
chr1	160112448	160112448	A	G	1	0	A85	ATP1A2	SI	-
chr1	160112523	160112523	A	G	0	1	A71	ATP1A2	SI	-
chr3	66429461	66429461	G	A	1	0	A25	LRIG1	SI	rs115135892
chr8	82353756	82353767	TCTATC TATCTA		1	0	A3	PMP2	NO	-
chr9	87641602	87641602	A	G	1	0	A10	NTRK2	SI	-
chr10	88810584	88810584	G	A	0	1	A67	GLUD1	SI	-
chr11	35272868	35272868	C	T	1	0	A92	SLC1A2	SI	rs147565042
chr11	35272878	35272878	T	C	1	0	A47	SLC1A2	SI	-
chr11	35273618	35273618	C	T	1	0	A3	SLC1A2	SI	rs187896453
chr11	35273673	35273673	G	A	0	1	A27	SLC1A2	SI	rs549790256
chr11	35273899	35273899	C	G	1	0	A88	SLC1A2	SI	rs144973554
chr11	35274486	35274486	A	C	1	0	A92	SLC1A2	SI	-
chr17	42983086	42983086	G	A	1	0	A55	GFAP	NO	rs567636153
chr17	42983430	42983430	G	A	1	0	A68	GFAP	NO	rs111477608
chr17	42992715	42992715	G	A	1	0	A101	GFAP	NO	rs57474185
chr19	45411110	45411110	T	C	1	0	A81	APOE	SI	rs769452

IV. 5. Estudio de la influencia de las variantes genéticas en la expresión. La variación en genes sobreexpresados explica diferencias de expresión entre casos y controles

A raíz de lo reportado en los apartados anteriores, los genes de módulos subexpresados en esquizofrenia son una mayoría del total de genes expresados que portan mutaciones (**tabla 4.15**), lo que explica que el análisis de expresión del total de genes mutados refleje que éstos se encuentren subexpresados en esquizofrenia en las muestras estudiadas (**figura 4.28**)

Con el propósito de estudiar la posible influencia de las variantes descritas en las diferencias de expresión génica observadas en esquizofrenia, se realizó un análisis de varianza (**métodos**) descomponiendo los valores de expresión génica de cada individuo en tres variables: fenotipo (caso o control), gen (cada uno de los genes presentes) y variante (presencia o ausencia de variante). De este modo, se buscó determinar la influencia de las variantes en la explicación de la varianza de la expresión génica a través de los distintos individuos mediante un análisis de la varianza (ANOVA de tres factores). La idea que subyace a este análisis es que, dado un diseño experimental apropiado, la variabilidad en la cantidad que se mide (la expresión génica en RPKM) se puede dividir en varias fuentes fácilmente identificables, mediante la comparación de los valores entre grupos y dentro de los grupos.

En este sentido, se analizó la variabilidad de la expresión génica en los tres grupos de genes considerados a lo largo del estudio de la variación genética (**tabla 4.18-A**): todos los genes mutados (623 *loci*), genes mutados de los módulos subexpresados en esquizofrenia M1 y M2 (345 *loci*) y genes mutados de los módulos sobreexpresados en esquizofrenia M10 y C38 (11 *loci*).

El análisis de la varianza reveló que la variable fenotipo contribuye de forma muy significativa a la explicación de la variabilidad en los valores de expresión génica a lo largo de los diferentes individuos, algo que es esperable en los módulos coexpresados e incluso en la totalidad de los genes mutados teniendo en cuenta que estos están enriquecidos en los genes subexpresados en esquizofrenia. Igualmente, la variabilidad en la expresión génica se explica de forma muy significativa por la variable gen, algo que es obvio dados diferentes niveles de expresión de los distintos genes en función de la naturaleza intrínseca de los mismos.

No obstante, también se observó una contribución significativa de la presencia o ausencia de variantes en los individuos para la explicación de la varianza en los datos de expresión, en los tres grupos considerados: todos los genes mutados ($p = 1,34 \times 10^{-12}$), genes de módulos subexpresados ($p = 0,00644$) y genes de módulos sobreexpresados ($p = 6,84 \times 10^{-3}$). Estos resultados demuestran la importancia de la variable "ausencia/presencia de variante" en los datos de expresión. Sin embargo, esta significación podría ser causada por dos posibles

escenarios. Por un lado, las variantes raras estudiadas podrían contribuir de forma significativa debido a la alteración que provocarían en los mecanismos bioquímicos de la expresión génica, siendo esta la hipótesis causal. Pero por otra parte, podría tratarse de una asociación indirecta producto de un escenario en el que las variantes raras presentes en genes altamente expresados y de forma ubicua en la totalidad de las muestras predispusiesen en mayor medida al desarrollo de la esquizofrenia. En este contexto, se podría observar una asociación indirecta debida al sesgo inherente en la expresión génica.

Tabla 4. 18. Resultados del análisis de varianza (ANOVA) para las tres variables (gen, fenotipo y variante) cuya influencia en la expresión se desea conocer. En la tabla se describen los grados de libertad (df), los valores de la F de Snedecor y el valor P de significación de la comparación de varianza entre grupos y dentro de grupos para los casos analizados (En todos los 623 genes mutados, los 345 genes mutados subexpresados y los 11 genes mutados sobreexpresados en esquizofrenia) con los datos de expresión (RPKM) sin normalizar y normalizando por gen.

	Expresión génica (RPKM)	A) Datos sin normalizar			B) Datos normalizados por gen		
		Df	F	P (>F)	Df	F	P (>F)
Todos los genes mutados (623)	Variantes	1	63,89	$1,34 \times 10^{-12}$	1	1,88	0,1703
	Genes	622	0,84	0,9986	622	0	10
	Fenotipo	1	83,85	$< 10^{-16}$	1	579,06	$< 10^{-16}$
	Residuales	28078			28078		
Módulos M1-M2 (345)	Variantes	1	3,42	0,00644	1	1	0,359
	Genes	344	5350,6	$< 10^{-16}$	344	0	1
	Fenotipo	1	174,08	$< 10^{-16}$	1	11800199	$< 10^{-16}$
	Residuales	15523			15523		
Módulos M10-C38 (11)	Variantes	1	20,68	$6,84 \times 10^{-3}$	1	6,51	0,01103
	Genes	10	129,18	$< 10^{-16}$	10	0,0172	1
	Fenotipo	1	32,14	$2,4 \times 10^{-5}$	1	72,71	$< 10^{-16}$
	Residuales	493			493		

A fin de contrastar estas dos hipótesis, se repitió el análisis de ANOVA para los tres grupos de genes pero con los valores de expresión normalizados por gen, eliminado así las diferencias intrínsecas en la expresión entre genes pero manteniendo las existentes dentro de un mismo gen (**tabla 4.18-B**). Bajo este nuevo contexto, las contribuciones por fenotipo siguen siendo altamente significativas, como es esperado, pero las contribuciones de las variantes en la totalidad de los genes mutados ($p = 0,17$) y en aquellos genes mutados de módulos subexpresados ($p = 0,359$) desaparecen, manteniéndose únicamente las correspondientes a los genes mutados pertenecientes a módulos sobreexpresados en nuestro estudio ($p = 0,011$),

justamente aquellos genes que demostraban una mayor presencia de variantes genéticas en casos respecto a controles (**tabla 4.14-B**).

Estos resultados demuestran que la regulación positiva en la expresión génica de los módulos enriquecidos en genes de astrocitos se puede explicar, al menos en parte, por la presencia de variantes en esos genes en las muestras, hecho que concuerda con la mayor proporción de variantes en casos respecto a controles (**tabla 4.14-B**). La totalidad de los valores de expresión en las muestras analizadas para los genes mutados de módulos sobreexpresados en esquizofrenia se muestran en el **ANEXO XXXXI- XXXXII**.







5. DISCUSIÓN



5.1. ESTUDIO I. VARIACIÓN GENÉTICA RARA EN ELEMENTOS CONSERVADOS DE *TCF4* DE POTENCIAL REGULADOR

En este primer estudio, se analizó la incidencia de variación genética rara en el elemento ultraconservado UC435, situado en la región intrónica de *TCF4*, en la susceptibilidad al desarrollo de esquizofrenia.

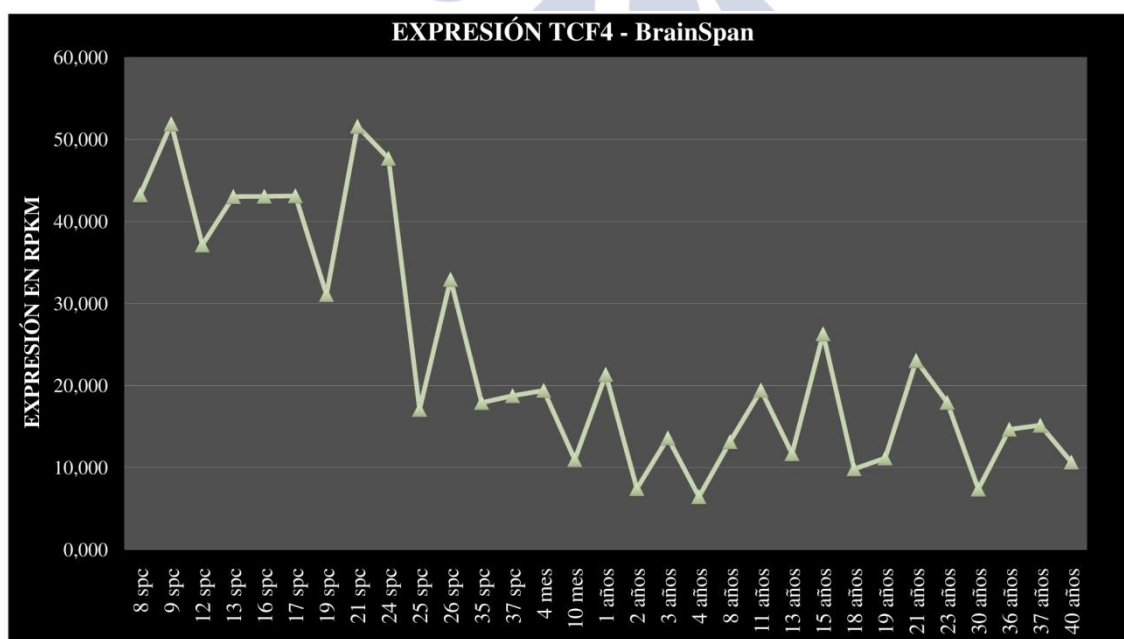
TCF4 (Entrez Gene ID 6925; ensembl ENSG00000196628), del inglés *transcription factor 4*, es un gen situado en el locus 18q21.2 que codifica para un factor de transcripción con una estructura básica hélice-giro-hélice (bHLH) y perteneciente al grupo de proteínas E, que se caracterizan por acoplarse entre ellas o con otras de estructuras similares para, posteriormente, unirse a la secuencia específica de ADN conocida como caja E (Ephrussi -box), regulando muchos procesos transcripcionales, sobretodo, durante las primeras fases del desarrollo. Este gen presenta una gran longitud en el genoma, abarcando hasta 437 kb y un total de 41 exones. A pesar de que *TCF4* se localiza en cerebro adulto, músculo, pulmones y corazón, esta proteína se encuentra, también, activamente expresada en diversos tejidos antes del nacimiento, jugando un papel en la maduración de las células para llevar a cabo funciones específicas como la diferenciación celular o la autodestrucción de las células (apoptosis). En cerebro, de hecho, *TCF4* se expresa consistentemente desde la concepción hasta la adultez, pero su picos máximos de expresión tienen lugar durante el neurodesarrollo, tal y como se deduce de los niveles de expresión recogidos de la base de datos de BrainSpan (**figura 5.1**). Esto ya viene a sugerir la clara influencia que ejerce este gen en la regulación de la expresión génica en procesos del neurodesarrollo. Estudios más pormenorizados de la expresión de *TCF4* demostraron que éste se localiza particularmente en áreas de alta plasticidad neuronal, tales como la corteza cerebral, hipocampo y cerebelo (Brzózka y cols., 2010). Por otro lado, también se ha descrito su expresión en el prosencéfalo y la zona ventricular del SNC embrionario (de Pontual y cols., 2009), así como en el telencéfalo en todas las etapas del desarrollo fetal (<http://www.brainspan.org/>). *TCF4* es, de hecho, la única proteína E que se expresa absolutamente en todas etapas del desarrollo del cerebro en ratón y humano (Quednow y cols., 2014).

La relación más clara entre variación genética en el gen *TCF4* y el sistema nervioso central proviene de los estudios que relacionan a este gen con el síndrome de Pitt-Hopkins (Pitt y Hopkins, 1978), una encefalopatía rara autosómica dominante causada en gran medida por mutaciones *de novo* no sinónimas y microdeleciones que afectan, principalmente, a la estructura de la hélice bHLH en la proteína en cuestión (Amiel y cols., 2007; Whalen y cols., 2012). La haploinsuficiencia de este gen durante el neurodesarrollo ha demostrado tener consecuencias muy severas que caracterizan el fenotipo de esta enfermedad (Zweier y cols., 2007), como un

grave retraso psicomotor, epilepsia, ataques diarios de hiperventilación, microcefalia o retraso mental (Peippo y cols., 2006; de Pontual y cols., 2009). Por este motivo, la expresión de este gen en procesos críticos del desarrollo prenatal debe estar severamente regulada para, con ello, evitar alteraciones en el normal desarrollo del cerebro.

En el contexto de la esquizofrenia, este gen también ha demostrado ser un factor clave desde el punto de vista genético al presentar variantes comunes de susceptibilidad al desarrollo de este trastorno ya en los primeros estudios de GWAS relevantes en esquizofrenia (Stefanson y cols., 2009; Steinberg y cols., 2011), cuando únicamente destacaban un número reducido de genes con significación genómica y entre los cuales se encontraba *TCF4*. Igualmente, la asociación se ha mantenido en los recientes GWAS de mayor potencia (Ripke y cols., 2013a; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

Figura 5. 1. Niveles de expresión (RPKM) del gen *TCF4* en cerebro humano desde las primeras etapas del neurodesarrollo hasta la adultez. Los datos de expresión reflejados en la figura se obtuvieron del promedio de los niveles de expresión del conjunto de tejidos disponibles en BrainSpan para cada etapa del desarrollo. Spc = semana post concepción; mes = mes desde nacimiento; años = años desde nacimiento.

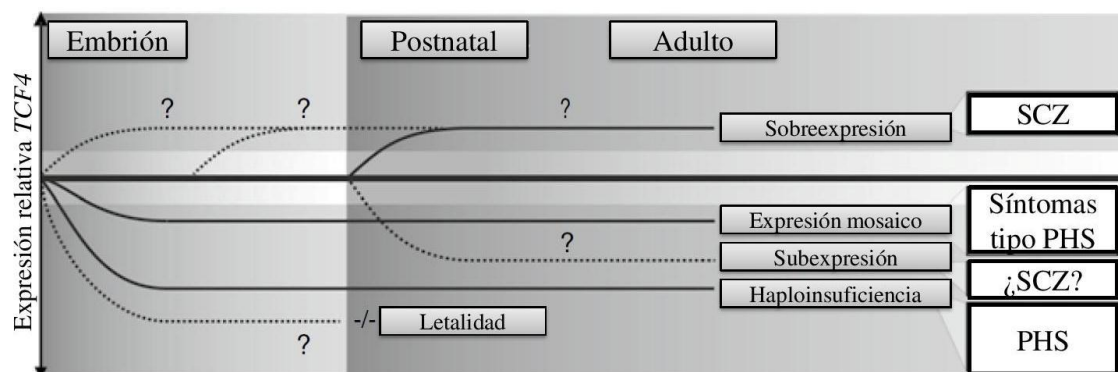


Sin embargo, a diferencia de lo observado en el síndrome de Pitt-Hopkins, las variantes de susceptibilidad en esquizofrenia no afectan a la estructura proteica, y se ubican en regiones no codificantes a lo largo de la secuencia del gen o alrededor del mismo. Por ejemplo, el estudio de Stefanson y cols., demostró la fuerte asociación del polimorfismo rs9960767 (OR = 1,23; valor $P = 4,13 \times 10^{-9}$), situado en el cuarto intrón de *TCF4*. Asimismo, el estudio de Steinberg y cols., 2011a, a parte del anterior SNP, reveló la existencia de asociación independiente con

rs4309482 (OR = 1,09; valor P = $7,8 \times 10^{-9}$), un SNP situado entre *TCF4* y *CCDC68*, 139 kb *downstream* del primero. Los intentos de buscar variantes no sinónimas que afectasen a la estructura proteica de TCF4, en DL con las de GWAS, no fueron exitosos (Williams y cols., 2011a). Sí lo fueron, sin embargo, otros estudios posteriores que revelaron la implicación del alelo de riesgo C de rs9960767 en varios endofenotipos de la esquizofrenia como la supresión sensorimotora de estímulos auditivos mediante medidas de la inhibición prepulso (Quednow y cols., 2011), la memoria verbal (Lennertz y cols., 2011) o el rendimiento cognitivo (Albanna y cols., 2014).

Estas asociaciones en regiones no codificantes de *TCF4* son coherentes con la existencia de un fenotipo esquizofrénico, menos severo que el del síndrome de Pitt-Hopkins, pero cuyo origen se fragua en el neurodesarrollo (Rapoport y cols., 2012), donde la expresión de este gen jugaría un papel clave. Se ha demostrado una sobreexpresión de *TCF4* en pacientes esquizofrénicos respecto a controles sanos tanto en sangre periférica (Wirgenes y cols., 2012) como en muestras de corteza prefrontal de enfermos (Guella y cols., 2013). Igualmente, neuronas originadas a partir de células pluripotentes inducidas de pacientes esquizofrénicos presentaron niveles incrementados de este gen (Brennand y cols., 2011). Estas observaciones, en general, apuntan a una alteración de la expresión de TCF4 en el ámbito de la esquizofrenia (Quednow y cols., 2014), pero en sentido opuesto al observado en el síndrome de Pitt-Hopkins, quedando por desvelar si la reducción en la expresión también puede causar síntomas esquizofrénicos, tal y como se ha observado de forma inconsistente en estudios en ratones (figura 5.2). Es lógico pensar, en consecuencia, que la implicación de *TCF4* en el desarrollo de la esquizofrenia podría estar mediada por variantes genéticas en regiones reguladoras que alterasen la correcta expresión del mismo, sobretudo en momentos críticos del neurodesarrollo.

Figura 5. 2. Relación entre los cambios en la expresión génica de TCF4 desde la etapa embrionaria hasta la adultez y los síntomas observados en humanos. PHS = Síndrome de Pitt-Hopkins; SCZ = Esquizofrenia. Figura modificada de Qednow y cols. (2014)



En este sentido, además, el gen *TCF4* constituye un factor de transcripción que actúa como pieza clave en el interactoma de genes de riesgo en esta enfermedad. Tal y como se confirmó en recientes estudios, su expresión está a su vez regulada por el *miRNA137*, un gen fuertemente asociado a esquizofrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) y que, a su vez, también interviene en la regulación postranscripcional de otros genes asociados (Collins y cols., 2014), como es el caso de *CACNA1C*, *CSMD1* (Kwon y cols., 2012) o *ZNF804A* (Kim y cols., 2012). Mientras, por otro lado, *TCF4* ha demostrado regular la expresión, mediante ensayos *in vitro*, de *NRXN1* y *CNTNAP2* (Forrest y cols., 2012), otros dos genes implicados en esquizofrenia principalmente a raíz del estudio de CNV de riesgo (Rees y cols., 2013).

Teniendo en cuenta que no existen alelos comunes con efectos moderadamente grandes sobre la expresión de este gen (> 20% en la diferencia), a partir de mediciones en cerebro adulto, que expliquen la sobreexpresión de *TCF4* observada en pacientes (Williams y cols., 2011a), es necesaria la búsqueda de variantes genéticas raras de susceptibilidad a esquizofrenia que pudieran interferir en la regulación de la expresión génica de este factor de transcripción. En este sentido, eventos mutacionales *de novo* en casos esporádicos de esquizofrenia han demostrado ser factores genéticos de susceptibilidad en aquellos genes particularmente importantes en el desarrollo neuronal prenatal (Xu y cols., 2012a).

El análisis de la secuencia del gen *TCF4* revela que no solo el motivo bHLH se encuentra altamente conservado, sino que el patrón de conservación genética que caracteriza a este gen es extensible a la totalidad de su secuencia, incluyendo la intrónica, y que se podría explicar mediante la necesaria selección purificadora en unas secuencias reguladoras importantes de cara a preservar su crítico papel en el desarrollo. Una muestra más de la versatilidad e importancia de este gen es la presencia de al menos 18 isoformas descritas (Sepp y cols., 2011), que sugieren la posible necesidad, también, de una correcta regulación del *splicing* alternativo de las mismas, en función del tejido y del momento del desarrollo. De hecho, de los 481 elementos ultraconservados descritos (Bejerano y cols., 2004), 2 de ellos existen en *TCF4* (UC435 y UC436) y uno de ellos es enteramente intrónico (UC435), por lo que fue seleccionado en el presente estudio para la búsqueda de variantes genéticas raras de susceptibilidad a esquizofrenia, en un contexto donde la alteración de la expresión génica de *TCF4* es un factor de predisposición.

Los UCEs son elementos de 200 pb como mínimo en donde existe una total homología de la secuencia genética comparando los genomas de rata, ratón y humano, y que predominan en regiones no exónicas de genes que regulan la transcripción de otros genes en el desarrollo precoz de la vida humana, así como en procesos del neurodesarrollo, tal y como se detalla en la

introducción (**apartado 1.3.3.2**). La mayor evidencia del papel regulador de la transcripción de estos elementos se obtuvo mediante ensayos *in vivo* de su capacidad de actuación como potenciadores o *enhancers* de la transcripción. De los 167 elementos ultraconservados probados en ensayos de gen *reporter* en pez cebra, el 45% actuaron como tales y, además, lo hicieron de forma particular en tejido cerebral y del tubo neural durante el desarrollo (Pennacchio y cols., 2006). Junto a otros trabajos, el de Penachio y cols. constituye la base de datos del buscador *VISTA enhancer* (Visel y cols., 2007; Visel y cols., 2009), donde se almacenan los elementos no codificantes con función de *enhancer* validados experimentalmente. El elemento UC435 es uno de ellos, demostrando la importancia funcional de esta región reguladora en la expresión de *TCF4*, y la potencial susceptibilidad al desarrollo de esquizofrenia que provocaría una alteración en la secuencia del mismo.

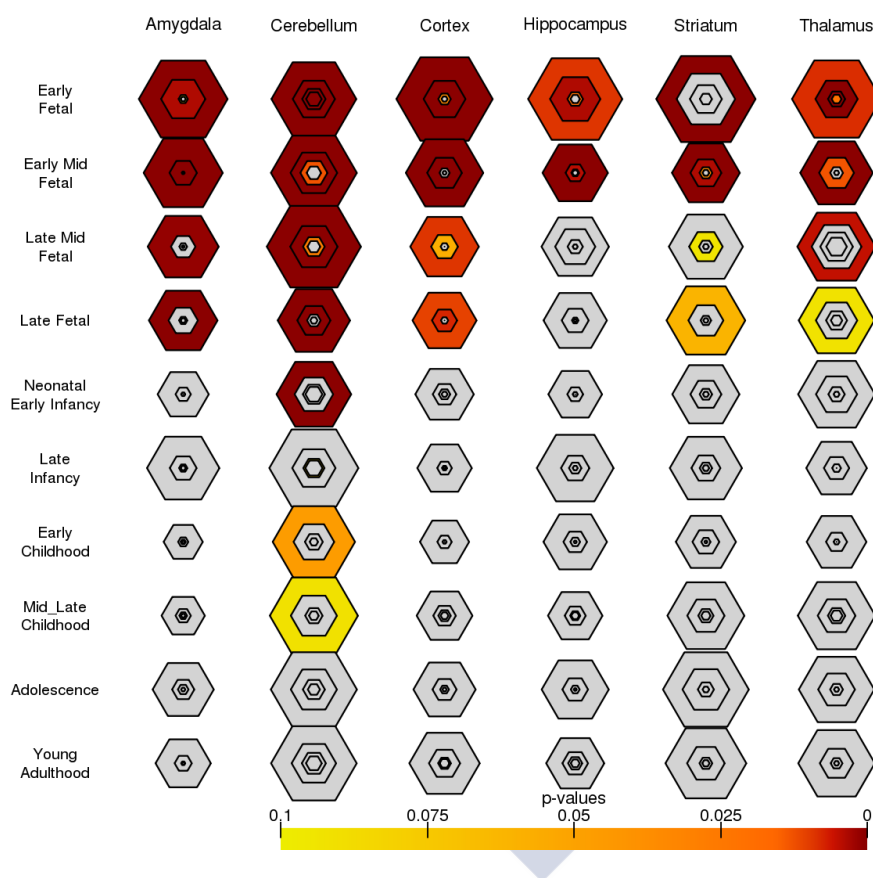
La comparación de los elementos UCE que afectan a regiones codificantes y aquellos que no en datos de expresión de BrainSpan a lo largo del desarrollo mediante la herramienta CSEA revela que los segundos se hallan fuertemente enriquecidos en los genes que se expresan en etapas tempranas del neurodesarrollo en la totalidad de las áreas cerebrales, demostrando su ubicuidad en procesos primordiales del desarrollo cerebral (**figura 5.3**). Sin embargo, en el caso de los genes que presentan UCEs en regiones codificantes, no existe tal enriquecimiento en ninguna región cerebral durante las distintas etapas del desarrollo. Por ello, esto es coherente con la idea de que la variación genética en estas regiones exónicas propiciaría un fenotipo mucho más extremo que el de la esquizofrenia, donde es más factible su atribución, en gran medida, a alteraciones en la expresión de genes de riesgo en algún momento concreto del desarrollo.

Dada la enorme conservación de esta secuencia analizada, se hipotetizó con la presencia de alguna variante rara pero, en ningún caso, de numerosos cambios nucleotídicos en la región UC435. Por ello, se pensó en la técnica de HRMA como una alternativa idónea para el presente estudio (**apartado 3.4**), tanto en el aspecto práctico como económico, dejando la técnica de secuenciación para la confirmación de las variantes detectadas en un barrido realizado con la primera.

La técnica HRMA, mediante el análisis de la pérdida de fluorescencia diferencial en variantes en heterocigosis, ha demostrado una técnica adecuada y empleada recientemente en el análisis de genotipado (Abidin y cols., 2015; Chen y cols., 2015; Lin y cols., 2014) y para la detección de mutaciones raras (Erali y cols., 2010). Para este último caso, la técnica de HRMA ha demostrado ser altamente eficaz (Vossen y cols., 2009). La detección de variantes en amplicones de hasta 4 kb ha demostrado una sensibilidad del 96% y una especificidad del 99,4%; llegando casi al 100% en amplicones de menos de 400 pb (Reed y cols., 2004). Además,

actualmente se considera uno de los mejores métodos para la detección de mutaciones raras dado que no es necesaria ninguna preparación de muestra ni separación y la relación de coste y efectividad es máxima, pudiendo incluso, recuperarse la muestra de trabajo una vez concluido el análisis, al no ser éste destructivo (Reed y cols., 2007).

Figura 5. 3. Análisis de CSEA (Xu y cols., 2014) para el enriquecimiento de los genes con UCEs en regiones intrónicas o intergénicas a menos de 200kb de distancia y las regiones del cerebro durante el desarrollo. En el ANEXO L se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.



En estudios de esquizofrenia concretamente, la técnica de HRMA ha sido igualmente empleada en el análisis de mutaciones a lo largo de la secuencia de genes candidatos como *DISC1* (Long y cols., 2013) o *SCN2A* (Woolf y cols., 2013).

Por todas estas razones, se buscaron variantes raras de susceptibilidad en la región ultraconservada UC435, y de función *enhancer*, en el gen *TCF4*, consistentemente asociado a esquizofrenia en numerosos estudios y factor clave en el neurodesarrollo. Tras los análisis en HRMA, ninguna muestra presentó variante a lo largo de la región UC435, y aquellas dudosas en su interpretación más estricta fueron secuenciadas, obteniéndose el mismo resultado negativo.

Con el análisis de 528 casos esquizofrénicos y 270 controles en el presente estudio (no se analizaron más controles al no encontrarse ninguna variante en casos), contamos con una probabilidad mayor al 80% para la detección de variantes genéticas con una frecuencia poblacional superior al 0,15% en la muestra de casos. Por tanto, podemos descartar la presencia de variantes alélicas en UC435 con una combinación de frecuencia y efectos similares a muchas CNV que confieren un riesgo moderado a la esquizofrenia. Esta equiparación no es, en absoluto, descabellada, teniendo en cuenta la enorme funcionalidad de las regiones ultraconservadas y, en particular, del gen *TCF4* en el neurodesarrollo. Por ejemplo, la duplicación en 16p11.2, y las deleciones en 1q21.1, 15q13.3 o 22q11.2, presentan una frecuencia superior al 0,15% en casos, pero claramente inferior a 0,1% en controles sanos (Malhotra y Sebat, 2012).

Una explicación a la ausencia de variantes, al margen de la potencia estadística, es precisamente la posibilidad de que la variación genética presente en esta región provoque un fenotipo más extremo que el esperado en esquizofrenia. Teniendo en cuenta la naturaleza del gen y la relevancia funcional de los elementos ultraconservados, una pérdida de función del mismo podría resultar en un cuadro clínico más severo. El hecho de que, tras extender los análisis de conservación de la secuencia de UC435 a 22 mamíferos placentarios adicionales y tres marsupiales, solo se hayan detectado dos cambios (uno en la posición 1 del elemento en hámster chino y otro en diablo de Tasmania), refuerza esta hipótesis. Sin embargo, contrariamente a dicha hipótesis, sorprende la presencia de tres variantes en población humana de los 1000 genomas, aunque en todos estos casos a una frecuencia menor al 0,1%.

Recientemente, un estudio de secuenciación del gen *TCF4* tampoco ha podido identificar variantes de riesgo a esquizofrenia pese a genotipar las variantes raras descritas en 1808 casos y 2261 controles (Basmanav y cols., 2015). Por todo esto, es evidente que son necesarios nuevos análisis con muestras de mayor tamaño para la búsqueda de variantes a frecuencia más baja que causen un efecto más agresivo en el fenotipo esquizofrénico observado para, con ello, probar si la hipótesis variante rara - enfermedad común es aplicable a la variación genética contenida en este elemento regulador altamente conservado.

5.2. ESTUDIO II. VARIACIÓN GENÉTICA EN ELEMENTOS ACELERADOS EN HUMANOS EN *NPAS3*.

En este segundo trabajo, teniendo en cuenta la evolución excepcional de las 14 regiones reguladoras aceleradas en su secuencia genética en el linaje humano (HAEs) a lo largo del gen *NPAS3* (Kamm y cols., 2013a), así como la importancia en el control de la expresión génica de *NPAS3* que presentan estos elementos, se estudió el impacto que podrían tener las variantes genéticas presentes en los elementos HAEs en el riesgo de aparición de esquizofrenia.

NPAS3 es un factor de transcripción que pertenece a la familia de proteínas de dominio estructural y funcional PAS. Estos dominios se encuentran en un gran número de proteínas presentes en casi todos los organismos, desde bacterias hasta humanos, y comparten importantes características estructurales y funcionales (Taylor y cols., 1999), entre las que destaca su actuación en la transducción de señales biológicas en base a un enorme rango de señales ambientales (Grossman y cols., 2001; Gu y cols., 2000; Repik y cols., 2000), como por ejemplo, activando la expresión génica en condiciones de hipoxia (Makino y cols., 2001). El dominio PAS fue nombrado a partir de las tres proteínas en la que fue descubierto por primera vez (Per, de *period circadian protein*; Arnt, de *Aryl hydrocarbon receptor nuclear translocator protein*; y Sim, de *single-minded protein*).

Las proteínas PAS más estudiadas son las que, además, presentan una estructura "básica" hélice-giro-hélice (bHLH-PAS), desde la que actúan como factores de transcripción gracias a su dominio de unión a ADN. Uno de los ejemplos más característicos de estas proteínas son las NPAS, proteínas PAS neuronales (Zhou y cols., 1997). *NPAS2* fue la primera proteína PAS descubierta en eucariotas (Dioum y cols., 2002), se expresa principalmente en cerebro anterior de mamíferos y está implicada en el control del ritmo circadiano (Hogenesch y cols., 1998, Reik y cols., 2002). *NPAS3*, por su parte, contiene 11 exones a lo largo de su estructura que codifican 3 dominios funcionales: una región de unión a ADN, una de dimerización (característica de los factores de transcripción de la familia bHLH que actúan conjuntamente regulando la expresión en forma de dímeros), y un dominio de transactivación, al que otras proteínas (como cofactores y otras proteínas) se unen para regular el proceso de transcripción (Gilles-Gonzalez y Gonzalez, 2005).

La translocación balanceada en 14q13 (Kamnasaran y cols., 2003; Pickard y cols., 2005) constituye la primera evidencia genética de la implicación del gen *NPAS3* en el desarrollo de la esquizofrenia (Kamnasaran y cols., 2003). En este estudio se demostraba que el papel de *NPAS3* como factor de transcripción de la familia *bHLH-PAS* se veía alterado en una madre y su hija, ambas con esquizofrenia y portadoras de la traslocación, dando lugar a una proteína truncada que, a pesar de que mantenía su capacidad de unión a ADN, carecía de los dominios de

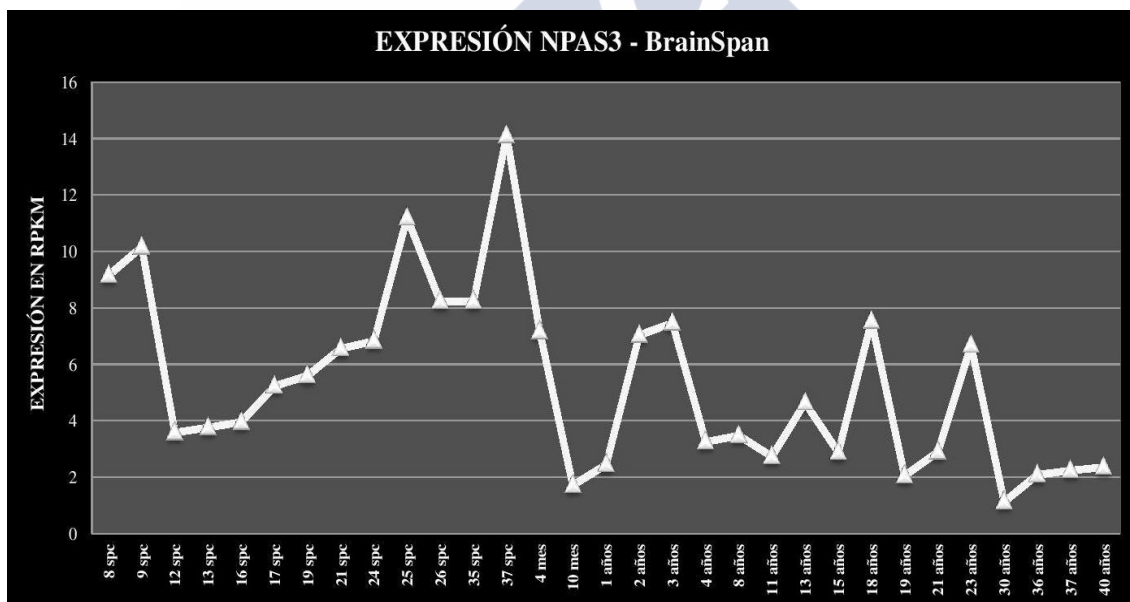
dimerización y de transactivación. Debido a esto, se postuló la haploinsuficiencia por pérdida de función de la proteína como un mecanismo por el que la variación genética en *NPAS3* aumentaba el riesgo de sufrir la enfermedad mental. Desde entonces, *NPAS3* ha sido objeto de estudios genéticos que han servido para describir la presencia de variantes raras y comunes en el gen asociadas con la susceptibilidad a la esquizofrenia y otros trastornos mentales (Pickard y cols., 2009; MacIntyre y cols. 2010; Yu y cols., 2014), aunque la búsqueda de variantes comunes de susceptibilidad a los trastornos psiquiátricos mediante estudios GWAS no han conseguido alcanzar significación genómica en ningún caso (Ferreira y cols., 2008, Huang y cols. 2010; Weber y cols., 2011). A pesar de ello, es digna de mención la significación de $p = 4 \times 10^{-6}$ alcanzada en el estudio de Huang y cols., que sugiere un estudio más pormenorizado de este gen en relación a los trastornos psiquiátricos y del neurodesarrollo.

Las asociaciones referidas se vieron reforzadas en estudios funcionales y de expresión. En esta línea, ensayos de *knock-out* de este gen, y de su relacionado *NPAS1*, en ratones, revelaron la existencia de muchos fenotipos conductuales, neurológicos y anatómicos relacionados con la esquizofrenia, tales como el deterioro de la cognición social, disminución de la respuesta de sobresalto, deterioro de la memoria, o la ampliación de los ventrículos (Erber-Sieler y cols., 2004, Brunskill y cols., 2005). A nivel celular, los ratones *NPAS3*^{-/-} mostraron sustanciales reducciones en los niveles de glicoproteína relina de las interneuronas gabaérgicas a través de todo el SNC (Erber-Sieler y cols., 2004), lo que concuerda con las reducciones tanto en ARNm como en nivel de proteína de relina observadas anteriormente en pacientes esquizofrénicos (Guidotti y cols., 2000; Impagnatiello y cols., 1998). En ratones adultos, la proteína *NPAS3* está altamente expresada en neuronas GAD67, uno de los tipos de neuronas corticales que presentan un cuadro más patológico en la esquizofrenia (Fung y cols., 2010; Guidotti y cols., 2000).

Es importante, para determinar el posible papel de *NPAS3* en la esquizofrenia como trastorno del neurodesarrollo, conocer cómo cambia la expresión del gen durante el desarrollo cerebral humano. Estudios previos en ratón describieron altos niveles de ARNm de *NPAS3* en la corteza cerebral durante los primeros estadios postnatales (Brunskill y cols., 1999), que disminuían progresivamente en la edad adulta, y de una forma coordinada con la maduración tardía de interneuronas inhibitorias en humanos (Fung y cols., 2010). A partir de los datos de expresión en cerebro durante el desarrollo de Brainspan, actualmente disponibles, se comprueba que la expresión del gen es consistente a lo largo de todas las etapas del desarrollo y adultez (siempre RPKM > 1), pero con importantes picos de expresión durante los últimos meses de desarrollo fetal (**figura 5.4**). Recientemente, Wong y cols. demostraron la importancia de la regulación postranscripcional en los niveles de *NPAS3* en córtex prefrontal (Wong y cols., 2012). En este estudio, se revelaba una reducción de la proteína *NPAS3* en la corteza prefrontal

de mujeres esquizofrénicas, debido a una mayor expresión de un microRNA, *miRNA17*, que regula la expresión de *NPAS3* a nivel de traducción, y cuya sobreexpresión en esquizofrenia ya había sido descrita anteriormente (Santarelli y cols., 2011). De hecho, el estudio de Wong y cols. demostraba una relación inversa entre los niveles de ARNm de *NPAS3*, regulados a la baja tras el nacimiento y hasta los 5 años, y los niveles incrementados de proteína NPAS3, lo que refleja la complejidad de los mecanismos regulatorios que cooperan para mantener niveles adecuados de este factor de transcripción de importancia crítica para el neurodesarrollo. La desregulación de los niveles de miRNA en corteza cerebral ha sido, de hecho, descrita en numerosos trabajos en relación con esquizofrenia y trastorno bipolar (Delalle y cols., 2014). El trabajo de Wong y cols., además del que describe la translocación de pérdida de función (Kamnasaran y cols., 2003), conforman la hipótesis de la hipofunción de *NPAS3* como factor de riesgo para el desarrollo de la esquizofrenia (Wong y cols. 2012).

Figura 5. 4. Niveles de expresión (RPKM) del gen NPAS3 en cerebro humano desde las primeras etapas del neurodesarrollo hasta la adultez. Los datos de expresión reflejados en la figura se obtuvieron del promedio de los niveles de expresión del conjunto de tejidos disponibles en BrainSpan para cada etapa del desarrollo. Spc = semana post concepción; mes = mes desde nacimiento; años = años desde nacimiento



Por otra parte, el gen *NPAS3* constituye la región genómica con un mayor número de elementos reguladores acelerados específicos de humanos (HAEs). Es decir, regiones altamente conservadas durante la evolución de mamíferos, pero cuya secuencia acumula cambios en el linaje humano tras la escisión respecto al del chimpancé (Kamm y cols., 2013a). Hay un total de 14 HAEs en *NPAS3*, siendo con notable diferencia, el gen con un mayor número de estos elementos (tabla 5.1). Aunque el gen *NPAS3* es el caso más paradigmático en este sentido,

otros genes implicados en la susceptibilidad a trastornos psiquiátricos también presentan elementos similares en regiones reguladoras. De los descritos en la tabla 5.2, destacan varios genes en el contexto de las enfermedades psiquiátricas y del neurodesarrollo. *CNTNAP2*, por ejemplo, es también uno de los genes con mayor número de HAEs y cuya haploinsuficiencia es un factor de riesgo al desarrollo de esquizofrenia y otros trastornos psiquiátricos, tal y como se demostró a partir del estudio de su delección (Friedman y cols., 2008; Peñagarikano y cols., 2011), así como de las variantes descritas en regiones reguladoras de su expresión génica (Chiocchetti y cols., 2014). Otro de los genes, *FHIT*, que presenta 8 elementos HAEs en su secuencia, es uno de los genes más subexpresados en modelos de ratón con fenotipo maníaco (Saul y cols., 2012). Por su parte, *RBFOX1*, una proteína de unión a ARN y también con 8 HAEs, ha demostrado, recientemente, una gran importancia en la regulación postranscripcional de genes particularmente implicados en el desarrollo cortical y el autismo (Lee y cols., 2016), y su disrupción es causa de alteraciones en la neurogénesis durante el desarrollo de la corteza cerebral (Hamada y cols., 2015), reforzando las evidencias previas de su implicación en esquizofrenia (Xu y cols., 2008), autismo (Sebat y cols., 2007) u otras enfermedades psiquiátricas (Davis y cols., 2012; Elia y cols., 2012; Martin y cols., 2007). Un último ejemplo es el gen *WWOX*, también portador de 8 HAEs y cuyo fenotipo neurológico asociado a mutaciones en su secuencia incluye convulsiones, ataxia, retraso en el desarrollo o espasticidad de gravedad variable, teniendo una gran relevancia en el desarrollo primordial del cerebro (Tabarki y cols., 2014).

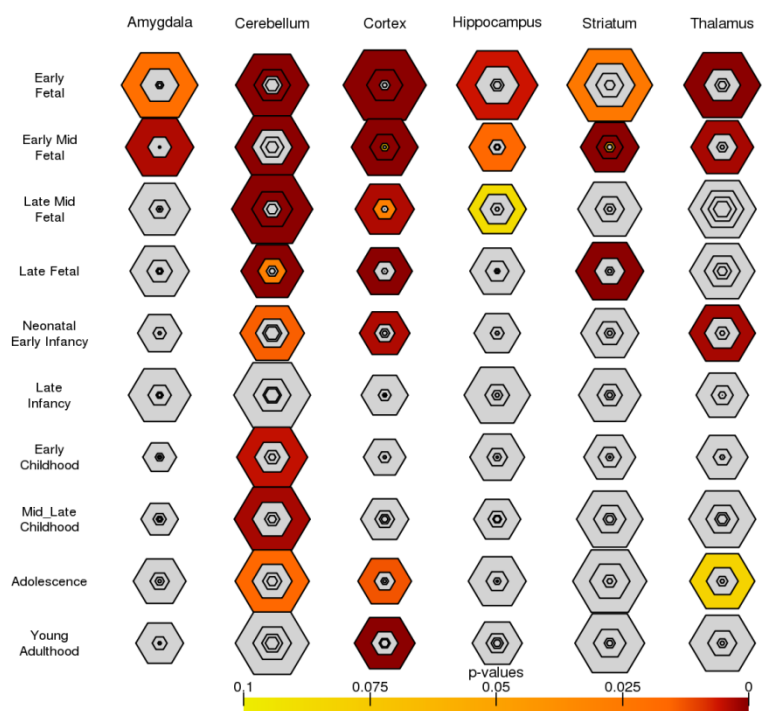
Tabla 5. 1. Genes con un número de elementos acelerados en humanos (HAEs) superior a 5. Tabla adaptada de Kamm y cols. (2013a).

Nº HAEs	Coordenadas (hg19)	Gen (RefSeq)	Nº HAEs/100 kb
14	chr14:33408459-34270951	<i>NPAS3</i>	1.62
8	chr3:59735036-61237133	<i>FHIT</i>	0.53
8	chr16:6069132-7763340	<i>RBFOX1</i>	0.47
8	chr16:78133551-79246564	<i>WWOX</i>	0.72
7	chr11:131240371-132206716	<i>NTM</i>	0.72
6	chr7:145813453-148118088	<i>CNTNAP2</i>	0.26
6	chr13:96743092-97491812	<i>HS6ST3</i>	0.8

Estos casos descritos sugieren la importancia de estos elementos acelerados en el linaje humano en un contexto de regulación de la expresión génica en procesos del neurodesarrollo críticos para el buen funcionamiento cerebral. Un análisis de CSEA, haciendo uso de los datos de BrainSpan, de los 616 genes en Gencode v.19 que solapan con elementos HAEs en su

secuencia revela que estas regiones se encuentran enriquecidas en genes particularmente expresados durante el neurodesarrollo (**figura 5.5**).

Figura 5. 5. Análisis de CSEA (Xu y cols., 2014) para el enriquecimiento de los genes presentes en Gencode v19 con HAEs en regiones las regiones del cerebro durante el desarrollo recogidas en Brainspan. En el ANEXO LII se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.



Curiosamente, este patrón de enriquecimiento es similar al de los genes con UCEs en sus regiones reguladoras (**figura 5.3**), tal y como sucede con *TCF4*, que presenta un elemento ultraconservado, UC435, exclusivamente en región no codificante (Bejerano y cols., 2004). No obstante, tal y como se estudia en la presente tesis (**apartado 5.1**), parece que no hay variantes poco comunes asociadas con el riesgo a esquizofrenia en esta región de *TCF4* (González-Peñas y cols., 2014).

Si la presencia de estos elementos sugiere un papel importante en la neurogénesis y desarrollo del cerebro humano en los genes que los poseen, el caso de *NPAS3* es todavía más destacable. Los 14 elementos HAEs que contiene se encuentran en su totalidad en regiones no codificantes del gen, mientras que la secuencia exónica está altamente conservada, tal y como lo estaban las primeras antes de sufrir aceleración en el linaje humano. Además, la función de los HAEs de *NPAS3* se ha descrito como reguladora de la expresión génica. En el trabajo de Kamm y cols. se demostró que 11 de los 14 HAEs actuaban como *enhancers* transcripcionales durante el desarrollo cerebral en los ensayos realizados de *gen reporter* en pez cebra (Kamm y cols., 2013a), algo que fue replicado posteriormente en ratón (Kamm y cols., 2013b). La actuación de

los elementos acelerados como potenciadores de la transcripción en otros de los genes que los contienen también fue consistentemente hallada en un estudio coetáneo (Capra y cols., 2013), en el que 24 de los 29 HAEs analizados regulaban la expresión en ratón.

La presencia de estos elementos acelerados preferentemente en regiones no codificantes de genes implicados en procesos de neurodesarrollo, sumado al hecho de su actuación como reguladores de la transcripción, sugiere la posibilidad de que sea la variación genética fijada por la selección en estas regiones, precisamente, la que provoca susceptibilidad al desarrollo de la enfermedad. En el ámbito de la esquizofrenia, una enfermedad propiamente humana con una heredabilidad del 64-80% y asociada a una fertilidad reducida, una hipótesis explicativa de su elevada prevalencia, cercana al 1% en todo el mundo, es la que defiende que varias mutaciones que han originado incrementos en habilidades cognitivas podrían estar favorecidas por selección positiva pero, a su vez, conferir mayor riesgo a esquizofrenia (Crow, 1997). En este sentido, diversos estudios han aportado evidencias empíricas de selección positiva en algunos genes asociados con la enfermedad, aunque es cierto que dichos estudios solo han centrado el foco en un número limitado de genes (Crespi y cols., 2007; Costas y cols., 2009; Lo y cols., 2007).

Dado que la descripción de elementos acelerados es relativamente reciente y que la mayoría de las variantes genéticas asociadas a esquizofrenia provienen de estudios de poco más de 2 años, pocos estudios han descrito el solapamiento de variantes genéticas en elementos acelerados en el linaje humano. Sin embargo, un estudio muy reciente (Xu y cols., 2015) demostró un alto enriquecimiento de los haplotipos de señales del GWAS más numeroso en esquizofrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) en genes separados 100 kb o menos de regiones conservadas en mamíferos y aceleradas en el linaje humano (mHARs) y en regiones conservadas en primates pero aceleradas en humanos (pHARs), que habían sido descritas previamente (Lindblad-Toh y cols., 2011). Además, los genes presumiblemente regulados por elementos conservados en primates y acelerados en humanos (genes p-HAR) se encuentran fuertemente conservados, lo que establece una diferencia clara entre la conservación de la estructura codificante de genes importantes para el normal desarrollo cerebral y las regiones que regulan su expresión (tal y como ocurre en *NPAS3*). Respecto a la totalidad de los genes asociados a esquizofrenia, aquellos que presentan elementos HARs cercanos (genes SCZ p-HAR) ocupan lugares más céntricos y conectados en los interactomas de los que forman parte y están estadísticamente enriquecidos en funciones de diferenciación neuronal u otras implicadas en el desarrollo cerebral y de la sinapsis, incluso cuando se comparan con la totalidad de genes asociados a esquizofrenia. De esta forma, estos genes representan un subconjunto "crítico" dentro de los genes asociados y, por ello, el análisis de HAEs representa una estrategia adecuada en la búsqueda de factores genéticos de mayor susceptibilidad. Una hipotética selección positiva ocurrida en estas regiones, de naturaleza

mayoritariamente reguladora, pudo haber sido clave en el salto evolutivo de las funciones cerebrales propias del linaje humano respecto al chimpancé, pero conllevar un riesgo asociado al desarrollo de trastornos psiquiátricos como la esquizofrenia.

Es importante señalar, de todos modos, que pese a que, en el estudio de Xu y cols., los genes cercanos a elementos acelerados en primates (genes PAR) no mostraban ninguna característica de las descritas en los genes pHAR ni los mHAR, otros estudios sí sugieren que ha habido más cambios en la secuencia y eventos potencialmente selectivos en el linaje del chimpancé que en el nuestro (Bakewell y cols., 2007; Hubisz y cols., 2014; Sequencing, T. C., & Analysis Consortium, 2005), por lo que no seríamos particularmente una especie con mayor aceleración en su secuencia genómica. Son necesarios, por tanto, más estudios que ayuden a encontrar una explicación para la aparición de estas regiones y la relación de la variación genética en ellas con los trastornos psiquiátricos y del neurodesarrollo.

En el presente estudio, se estudiaron 21 variantes de elementos HAEs presentes en población de los 1000 genomas, así como 5 SNP procedentes de estudios GWAS en NPAS3, no habiendo asociación alguna ni desde el punto de vista individual (**tabla 4.1**) ni conjuntamente (**tabla 4.2**). En este último caso, se realizaron diferentes ensayos en base a distintos modelos con los que ampliar los escenarios de estudio acerca del posible papel de variantes raras en comparación con las comunes, así como de los efectos de protección frente a los perjudiciales, por lo que la utilización de SKAT nos permite aseverar la ausencia de asociación en un conjunto amplio de posibilidades. De igual modo, el análisis de haplotipos, utilizando un procedimiento de "ventanas deslizantes" de 2 a 6 SNP, tampoco reveló ninguna asociación significativa tras corregir por las múltiples pruebas realizadas (**tabla 4.3**).

El análisis de potencia sugiere que el presente estudio tiene capacidad suficiente como para poder detectar asociación en diversos escenarios realistas, de acuerdo con el conocimiento actual de la arquitectura genética de la esquizofrenia (**figura 4.6**). Por ejemplo, se alcanzaría una potencia del 80% en el caso de que un 25% de las variantes estudiadas fuesen funcionales, y en las que el OR máximo de éstas fuese de 5, asignado a las variantes con MAF = 0,0001.

Este contexto es perfectamente plausible, tal y como se destacó anteriormente en el **estudio I** de TCF4 para el riesgo asociado de numerosas CNV en esquizofrenia (OR>5). Del mismo modo, Purcell y cols. (2014) secuenciaron los exomas de 2536 casos de esquizofrenia y 2543 controles, identificando tres pequeños conjuntos de genes (complejo de la proteína regulada por actividad ARC asociada al citoesqueleto, complejo de proteínas post-sinápticas PSD95 y genes de los canales iónicos de calcio dependientes de voltaje) en los que se acumulan variantes raras de gran efecto funcional (OR de 19,1, 5,1 y 8,4, respectivamente). Por tanto, la falta de resultados significativos en nuestro análisis sugiere que, contrariamente a la hipótesis

inicial formulada, la evolución especial acontecida en *NPAS3*, y reflejada en la existencia de 14 elementos HAEs acelerados en *Homo*, relajó la fuerte conservación característica de estos elementos a lo largo del linaje de mamíferos, sin implicaciones en el desarrollo de la esquizofrenia.

De hecho, el estudio de variación genética en regiones conservadas o aceleradas en el linaje humano también deja otros estudios previos con resultados negativos (Schanze y cols., 2012; Tolosa y cols., 2008), por lo que las apreciaciones al respecto de un salto evolutivo que tuviera como contrapartida la aparición de trastornos psiquiátricos deben ser tomadas con cautela, y a la espera de estudios de mayor potencia centrados en estas regiones. En este sentido, se ha descrito que muchos de los elementos acelerados se encuentran en regiones próximas a los extremos de los cromosomas (Pollard y cols., 2006; Prabhakar y cols., 2006), donde tiene lugar una tasa mayor de recombinación (Lander y cols., 2001), que se asocia a un sesgo en las variantes nucleotídicas de AT a CG (Duret y Galtier, 2009). 8 de los 14 HAEs presentan este sesgo en sus variaciones específicas en humanos (HAR189-HACNS489, HACNS553, HAR173, HACNS96, 2xHAR142, HAR89, 2xHAR223 y 2xHAR122). Aunque *NPAS3* no se encuentra situado cerca del extremo del cromosoma 14, no se puede descartar la posibilidad de que la conversión de AT a CG, tras una relajación de la conservación por una pérdida de la importancia funcional en el linaje humano, hubiera contribuido a la evolución particular del gen. Sin embargo, en un estudio desarrollado para estudiar los efectos de la selección y la conversión sesgada de AT a CG mostró que más de la mitad de los HAEs (55%) tienen tasas de sustitución que exceden significativamente las estimaciones en relación a la neutralidad, lo que sugiere que estas regiones pudieron haber sido originadas por selección positiva en lugar de por relajación de los patrones de conservación de los genes en los que se encuentran (Kostka y cols. 2012).

La comparación de los *loci* que albergan un número excepcional de HAEs con los datos del mayor estudio GWAS en esquizofrenia hasta la fecha (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) confirma la relevancia de la variación genética en estos elementos. El intervalo de 1 Mb genómico con mayor número de HAEs, que incluye el extremo 3' de *AKAP6* y parte de *NPAS3*, presenta un SNP cercano a la significación genómica (rs12883788, $P = 8,98 \times 10^{-8}$, incluyendo muestras de replicación). Este SNP se encuentra en el 3' UTR de *AKAP6*, 100kb *upstream* de *NPAS3*. Por otro lado, entre los siete genes con más de cinco HAEs, cuatro de ellos contienen SNP en rangos de significación de $P < 1 \times 10^{-5}$. El segundo gen por el número de HAEs, *FHIT*, contiene un SNP asociado a esquizofrenia significación genómica en la muestra de descubrimiento (rs1353545, $P = 4,26 \times 10^{-8}$, en 34241 casos y 45604 controles), aunque su valor de significación fue ligeramente menor en el conjunto total de datos ($P = 7,4 \times 10^{-8}$, tras la inclusión de 1513 nuevos casos y 66236 controles adicionales como muestras de replicación). Los otros tres genes son *NPAS3* (rs7140760, $P =$

$7,26 \times 10^{-6}$), *RBFox1* (rs12447542, $P = 1,12 \times 10^{-6}$), y *HS6ST3* (rs9554348, $P = 3,61 \times 10^{-6}$). No obstante, ninguno de los cinco SNP se encuentra dentro de los elementos HAEs y no hay variantes en HAEs en DL con estos SNP en muestras europeas de los 1000 genomas.

Por otra parte, una de las cinco variantes tomadas de GWAS independientes para este estudio, rs8017055, se asocia nominalmente a esquizofrenia ($P = 1,36 \times 10^{-4}$). Este SNP pertenece, de hecho, a la región que se asocia más significativamente dentro del gen *NPAS3*. rs8017055 es la variante genotipada más cercana al extremo 3' de *NPAS3* en el presente estudio, y 18,7 kb *upstream* del HAE 2xHAR122, no representado en este trabajo por no presentar ninguna variante en los 1000 genomas en muestras europeas pero que sí presenta hasta 7 variantes en otras poblaciones diferentes. De hecho, otros SNP asociados en el citado GWAS de esquizofrenia, y localizados en la región 3' de *NPAS3*, están todavía más cerca del elemento no representado 2xHAR122. El SNP más cercano, rs8009687 ($P = 1,39 \times 10^{-5}$), y el SNP más asociado, rs7140760 ($P = 7,27 \times 10^{-6}$), se encuentran a tan sólo 7,4 kb y 8,5 kb *upstream* de 2xHAR122, respectivamente. Estas asociaciones sugieren una posible implicación de este HAE en la esquizofrenia pero que no pudo ser probada mediante la metodología empleada en este estudio debido a la utilización de muestras de origen europeo. La ausencia de variantes en 2xHAR122 en la población de los 1000 genomas sugiere, igualmente, la restricción de esta región frente a variantes que pudieran alterar la funcionalidad alcanzada tras los cambios ocurridos en el linaje humano.

Este caso es extensible a otros 2 HAEs (HACNS553 y HAR173), que no presentan ninguna variación genética de acuerdo con la base de datos de población europea en los 1000 genomas pero sí en otras poblaciones, lo que hace un total de 3 de 14 HAEs que no están siendo considerados. Es por ello por lo que son necesarios estudios similares en otras poblaciones no europeas para extender este análisis a los 14 HAEs descritos en *NPAS3*. La resecuenciación de estos elementos acelerados y/o conservados (y otros presentes igualmente en genes relacionados con la función cerebral), en muestras de mayor tamaño, es un requisito indispensable para descartar la implicación de estos elementos en la esquizofrenia, buscando posibles variantes raras de efecto todavía mayor o variantes más frecuentes y con un riesgo asociado menor, objetivos que no son alcanzables desde el presente tamaño muestral.

5.3. ESTUDIO III. RESECUENCIACIÓN DE REGIONES PROMOTORAS DE GENES DE RIESGO A ESQUIZOFRENIA.

En este tercer trabajo del presente proyecto de tesis se analizaron 74 kb de secuencia reguladora pertenecientes a 29 regiones promotoras en 516 casos y 516 controles agrupados en sendos 83 *pools* de muestras, algo que fue posible gracias a una exhaustiva metodología que comprendió el diseño y amplificación de las regiones reguladoras mediante PCR con enzimas *High fidelity* y la posterior resecuenciación dirigida de estas regiones con la plataforma de ultrasecuenciación de SOLiD. El objetivo principal del presente trabajo fue la detección de variantes genéticas raras o infrecuentes, en coherencia con la hipótesis variante rara – enfermedad común sugerida, y por la que la mayor parte de las variantes de riesgo presentes en estas regiones, de elevada importancia reguladora, estarían sometidas a selección purificadora y no alcanzarían una frecuencia de polimorfismo (Manolio y cols., 2009; McClellan y cols., 2007). Por ello, en el proceso de detección se estableció un protocolo de filtros de calidad que asegurase la igual representatividad de todas las muestras (**figura 4.13**) y la eliminación de posibles artefactos de amplificación o secuenciación. Así, aún a costa de perder algunas posibles variantes reales, se preservaron aquellas de alta calidad y profundidad de lectura (**tabla 4.4, tabla 4.5**). La eficacia de este proceso fue, de hecho, comprobada mediante el genotipado de las variantes con predicción funcional más patogénica (**apartado 4.III.4**), algo que no es práctica habitual en muchos de los proyectos de secuenciación masiva realizados actualmente, y que sitúa los resultados aquí presentados en un nivel de confianza muy elevado. Siguiendo nuestra metodología, un total de 361 y 388 variantes en casos y controles, respectivamente, relativas a 297 posiciones mutadas fueron detectadas en estas 29 regiones, cuyo potencial regulador de la expresión génica también fue sugerido en este trabajo mediante el análisis *in silico* de predicción funcional (**figura 4.12**). El análisis de las mismas reveló un enriquecimiento de variantes en las posiciones de mayor evidencia reguladora en casos frente a controles y, de forma destacada, una implicación de las variantes que afectan a los estados represivos de la cromatina en la región 17q12 en la etiología de la esquizofrenia, tal y como se discute a continuación.

El estudio de las regiones reguladoras de forma sistemática en relación al riesgo que confiere la variación genética a trastornos complejos tiene sentido en un contexto como el actual para la esquizofrenia y otros trastornos psiquiátricos. Los estudios de variación común en los últimos GWAS han demostrado que las principales asociaciones son debidas a variantes en regiones no codificantes y reguladoras de la expresión génica (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Richards y cols., 2012; Roussos y cols., 2014). En este sentido, el estudio de Roussos y cols. reveló un enriquecimiento significativo de las variantes asociadas a esquizofrenia del último GWAS por aquel entonces (Ripke y cols., 2013a)

en elementos reguladores de la expresión en *cis* (CRE). En el mismo, se encontró un solapamiento significativo entre SNP que afectaban a la expresión (eSNP) y esos elementos CRE (creSNP), en consonancia con estudios previos que revelaban una asociación entre alelos de susceptibilidad y aquellos que afectan a la expresión en cerebro humano (Richards y cols., 2012). Roussos y cols. encontraron la mayor asociación con esquizofrenia en los creSNP de *enhancers* en cerebro fetal y adulto, confirmando hallazgos previos que ya señalaban la importancia crítica de los *enhancers* en la organización de la transcripción durante el neurodesarrollo (Wenger y cols., 2013). Estos autores también demostraron la interacción en *trans* entre elementos reguladores distales y proximales en el locus del gen *CACNA1C*, sugiriendo la existencia de redes de mecanismos bioquímicos de regulación de elevada complejidad.

No obstante, la escasa heredabilidad explicada por las asociaciones derivadas de estudios GWAS de gran potencia (Ripke y cols., 2013a; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), sumado a los pobres resultados de los estudios de secuenciación de exomas (Need y cols., 2012; Purcell y cols., 2014), sugiere la presencia de variantes genéticas de susceptibilidad de baja frecuencia localizadas en regiones no codificantes del genoma que, afectando a la expresión génica, contribuyen al riesgo a desarrollar esquizofrenia. En esta línea, el reciente estudio de Duan y cols., centrado en la búsqueda de variantes raras de predisposición a esquizofrenia que afectasen a la expresión del gen *MIR137*, uno de los genes más fuertemente asociados a esquizofrenia en los estudios GWAS pero sin resultados similares en los estudios de exoma, se basó en la secuenciación de casi 7 kb de regiones reguladoras alrededor de dicho gen en 2610 casos y 2611 controles (Duan y cols., 2014). Para la selección de regiones reguladoras, los autores se basaron en las regiones accesibles del gen, de hipersensibilidad a DNasa (DHS) de las células neuronales presentes en la base de datos de ENCODE (SK-N-SH y NH-A) y de cerebro fetal, anotando estas secuencias seleccionadas en función de las marcas de histonas (**tabla 1.8**) como promotores (presencia de H3K4me3), *enhancers* (presencia de la marca H3K4me1) o *insulators* (cuando presentaban la marca represiva de CTCF, que aísla el efecto del *enhancer*). El análisis filogenético de las regiones reguladoras seleccionadas reveló unos niveles de variación genética similares a los esperados para variantes *missense* o *frameshift*, por lo que claramente eran regiones sometidas a selección purificadora. Este hecho refuerza nuestras hipótesis, ya señaladas en la presente discusión, acerca de una selección negativa que relega a la mayor parte de la variación con consecuencias funcionales a frecuencias no detectables por los estudios de GWAS basados en DL.

Los resultados del estudio de Duan y cols., además, mostraron una acumulación significativa en casos esquizofrénicos respecto a controles sanos de variantes en regiones activas

de promotores y *enhancers*, mientras que la significación desaparecía si se incluían las variantes en *insulators* (enriquecidas, además, en controles aunque no significativamente). Estos datos, a parte de la obvia interpretación del riesgo que conlleva esta variación reguladora en regiones no codificantes del gen, revelaron la importancia de analizar la acumulación de variantes raras en estas regiones en función de los elementos funcionales afectados, ya que muchas veces pueden tener tendencias contrapuestas, como en el caso. Sin la información de la caracterización funcional necesaria para dividir elementos activos y represores de la transcripción (basada en la distinta combinación de marcas histónicas), no se habría podido desvelar el riesgo asociado específicamente a las variantes en las regiones reguladoras activas. El estudio permitió, además, tras la identificación de una variante concreta en 11 casos y 0 controles, y su posterior genotipado en miles de casos independientes, demostrar la asociación de la misma con la esquizofrenia, describir el impacto que provocaba en la unión a TF y su funcionalidad bioquímica como *enhancer* específico de tejido. Por tanto, este estudio representa un ejemplo de cómo algunos genes asociados a esquizofrenia, pero sin variaciones en su secuencia codificante, pueden ejercer su predisposición a través de variantes reguladoras, menos agresivas que las codificantes, y que concuerdan en mayor medida con un fenotipo común en la población general, de forma similar a lo que se señalaba en el primer bloque de la introducción donde se contraponía el efecto de las mutaciones no sinónimas de *TCF4* asociadas a un trastorno de fenotipo extremo como el PHS frente al de las variantes asociadas a esquizofrenia en el mismo gen a partir de estudios GWAS, de naturaleza no codificante.

La razón fundamental que dio lugar al proyecto ENCODE fue la descripción y caracterización funcional de las regiones no codificantes (ENCODE Project Consortium, 2012). En este sentido, fue de gran relevancia para el presente trabajo la demostración de que los promotores se extienden simétricamente a ambos lados del TSS abarcando un promedio de 2,5 kb (Birney y cols., 2007), y no únicamente *upstream* del TSS. Esta novedosa información fue asimilada en el trabajo aquí discutido, por lo que se amplificaron sistemáticamente las regiones reguladoras promotoras pertenecientes a genes candidatos en esquizofrenia en CNV y estudios de GWAS, siguiendo esas premisas mediante el procedimiento descrito (**métodos**). A pesar de la existencia de otras regiones alejadas del TSS de gran importancia para la regulación de la expresión génica, la selección de regiones promotoras representa una buena aproximación a la búsqueda de variantes reguladoras de susceptibilidad por su unívoca implicación en el control de la expresión del gen en que se localizan. En este sentido, las interacciones de elementos reguladores distales (como por ejemplo, *enhancers*) son más difíciles de identificar que las que implican elementos proximales (Gerstein y cols., 2012; Yip y cols., 2012). Por otro lado, los tagSNP en DL con las regiones 5'UTR, incluidas en gran medida en la nueva definición de promotores, son los más enriquecidos en *loci* de riesgo de esquizofrenia en comparación con

otras categorías, incluidas las variantes exónicas (Schork y cols., 2013), lo que vuelve a demostrar la gran implicación de estas regiones en el riesgo asociado a esta enfermedad.

En el trabajo aquí discutido se emplearon datos generados por el proyecto ENCODE (ENCODE Project Consortium, 2012; **tabla 1.7**), y recogidos en la base de datos de la herramienta RegulomeDB (Boyle y cols., 2012), para revelar un exceso de SNV raras en pacientes con esquizofrenia, a lo largo de las regiones secuenciadas, en aquellas posiciones de mayor predicción funcional. Esto es, para el caso, la calificación “2” de RegulomeDB (**tabla 1.10**), que implica la mayor evidencia de ser sitio de unión de factores de transcripción después de la categoría 1 (aquí no considerada por necesitar, además de los requisitos de la categoría 2, evidencias de eQTL, lo cual descarta de forma sistemática la inclusión de las variantes raras).

La importancia de la variación reguladora y la eficacia de la herramienta RegulomeDB en la detección de la misma ya fue demostrada al revelarse que un 80% de las asociaciones detectadas en estudios GWAS de enfermedades complejas pueden interpretarse funcionalmente mediante esta herramienta (Schaub y cols., 2012). En el contexto de la esquizofrenia y las enfermedades psiquiátricas, estas evidencias han llevado a que, actualmente empiece a emplearse la base de datos de RegulomeDB, que integra los datos de ENCODE, en la interpretación de variantes funcionales. Un reciente estudio que analiza la variación rara en 26 genes que codifican canales iónicos neuronales, incluyendo subunidades de los receptores GABA y los canales de calcio dependientes de voltaje, en trastorno bipolar, demostró que las variantes que afectan a sitios de unión de factor de transcripción según RegulomeDB situadas en promotores y *enhancers* estaban significativamente enriquecidas en casos (Ament y cols., 2015). Este hecho contrastaba, además, con la menor importancia de las variantes codificantes de los mismos genes asociados, algo comparable a lo ocurrido con el gen *MIR137* en el estudio de variantes reguladoras en esquizofrenia (Duan y cols., 2014). Igualmente, otros estudios han demostrado, mediante la utilización de los datos de ENCODE en RegulomeDB, la importancia de las asociaciones en regiones reguladoras. Tales trabajos, entre otros, describieron las asociaciones en esquizofrenia de variantes no codificantes en *CTCF* (Juraeva y cols., 2014), de variantes intrónicas en haplotipos del gen *MLK1* (Luo y cols., 2014) o de un haplotipo de resistencia a antipsicóticos formado por variantes que interrumpen la unión de factores de transcripción en el gen *PIP4K2A* (Kaur y cols., 2014).

Es interesante señalar que el riesgo asociado al grupo de variantes seleccionadas por ser RegulomeDB 2 (**figura 4.12**; OR = 1,92 (1,26 - 2,95) fue, además, superior al riesgo medio de las variantes exónicas (OR = 1,12) en el mayor estudio de secuenciación de exoma hasta la fecha (Purcell y cols., 2014), para variantes todavía más raras (MAF < 0,001 y MAF < 0,005) (**tabla 1.6**). Esta comparación sugiere una probable mayor contribución de las variantes raras no

codificantes de importancia reguladora, en enfermedades comunes como la esquizofrenia, respecto a las que alteran la secuencia aminoacídica, de forma análoga a lo ampliamente demostrado en variación común (Albert y Kruglyak., 2015; Cirulli y Goldstein, 2010; Gusev y cols., 2014; Maurano y cols., 2012; Roussos y cols., 2014).

Por otro lado, en una segunda parte del presente trabajo, se analizaron las anotaciones epigenéticas relativas a los dos tejidos más implicados en el desarrollo de la esquizofrenia según los resultados del análisis del último GWAS en esquizofrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), que situaban el mayor enriquecimiento de *loci* asociados a esquizofrenia en las regiones reguladoras de giro angular y lóbulo frontal medio, y las relativas a cerebro fetal, escogido siguiendo la hipótesis del neurodesarrollo, al escoger genes de CNV implicadas en este tipo de trastornos. Mediante este análisis, haciendo uso de los datos epigenéticos específicos para estos tejidos del proyecto Roadmap (Kundaje y cols., 2015), se observó un exceso de variantes raras en casos respecto a controles en aquellos estados represivos de la cromatina, y particularmente en aquellos de funcionalidad de represor Polycomb (**tabla 4.6**), caracterizados por la modificación histónica H3K27me3 y relacionados con genes homeóticos de importancia primordial en el desarrollo. En este sentido, la clasificación en 15 estados de la cromatina específicos de tejido propuesta por el proyecto Roadmap fue realizada en base a 5 modificaciones histónicas descritas de forma consistente para 127 tejidos (**tabla 1.9**), ampliando la información relativa a la importancia de las mismas ya destacada en los resultados del proyecto ENCODE (ENCODE Project Consortium, 2012). Fue, de hecho, en el proyecto ENCODE en el que se describió que las regiones reguladoras promotoras explicaban la mayor parte de la variabilidad en los niveles de expresión génica correlacionándose fuertemente con las marcas de acetilación y metilación en las histonas (r de Pearson = 0,9) y, en menor medida, con la unión de factores de transcripción (r de Pearson = 0,81) (Dong y cols., 2012; Sanyal y cols., 2012; Thurman y cols., 2012). En general, las modificaciones activadoras (H3K27ac, H3K9ac, H3K4me3 o H3K4me2) explican la mayor parte de esta variación, aunque para un grupo reducido de elementos promotores, las marcas represivas (H3K27me3 o H3K9me3) son particularmente importantes a la hora de predecir los niveles de expresión, siendo ambas altamente dinámicas a lo largo de la vida.

En el caso de nuestro estudio, el análisis inicial de enriquecimiento de las variantes en los distintos estados cromatínicos revela que nos encontramos ante la segunda situación, más infrecuente, pero coherente con el criterio de selección de genes implicados en el neurodesarrollo como los elegidos de estudios de CNV. Este análisis, reflejado en la **tabla 4.6**, pone de manifiesto que la división de las regiones secuenciadas en función de su estado epigenético es imprescindible para poder visualizar, de una forma más específica, la implicación de la variación reguladora en cada uno de los mismos, una estrategia de demostrada y destacada

utilidad precisamente en la búsqueda de variantes raras reguladoras en esquizofrenia (Duan y cols., 2014), pero que, con los actuales datos de Roadmap, permite un análisis de los estados epigenómicos mucho más pormenorizado.

Además, en el presente trabajo, la combinación de las dos fuentes de información utilizadas (datos de DHS y de TFBS de ENCODE según RegulomeDB y datos epigenómicos de Roadmap) dio lugar a una diferencia todavía más significativa. En concreto, nueve de los 516 pacientes esquizofrénicos y ninguno de los 516 controles existentes fueron portadores de variantes raras de potencial regulador de la expresión génica (RegulomeDB 2) y localizadas en regiones reprimidas de la cromatina en el estado represor Polycomb (**tabla 4.7**), una diferencia altamente significativa (P valor de prueba exacta de Fisher de dos colas= 0,0016, OR = 19,34, 95% C.I. = 2,45-2495,26). De hecho, el riesgo atribuible a estas variantes, a pesar del moderado número de individuos analizados, supera al de muchas CNV, sugiriendo una importancia funcional mayúscula de estas variantes en lo que se refiere al riesgo asociado a esquizofrenia. Igualmente, el efecto de estas variantes es equiparable al de aquellas codificantes y de pérdida de función en los sets de genes más asociados en secuenciación de exoma (Purcell y cols., 2014).

Estas 9 variantes localizadas en regiones con esta funcionalidad particular de represores Polycomb con una alta evidencia de su importancia reguladora, se corresponden con 6 posiciones situadas exclusivamente en los genes *HNFI1B* y *LHX1*, pertenecientes a la delección del locus 17q12. Esta CNV, además, se caracteriza por su muy significativa asociación no solo a esquizofrenia (OR = 6,24; P = 0,0072), sino a otros trastornos del neurodesarrollo como discapacidad intelectual (OR = 17,3; P = $1,4 \times 10^{-12}$; Malhotra y cols., 2012) o autismo (OR = 16; P = 0,01; Malhotra y cols., 2012), subrayando la implicación de la misma en el anormal neurodesarrollo cerebral. Los genes *HNFI1B* y *LHX1* codifican para factores de transcripción con estructura de homeodominio e implicados en el desarrollo de múltiples órganos, lo que explica que sus regiones promotoras se caractericen por el estado epigenético de represor Polycomb. *HNFI1B* juega un papel importante en la segmentación de parte posterior del cerebro de los vertebrados durante el desarrollo (Chomette y cols., 2006; Pouilhe y cols., 2007; Sturgeon y cols., 2011; Sun y cols., 2001), algo muy relevante teniendo en cuenta la implicación del cerebelo, una de las estructuras anatómicas surgida de la parte posterior del cerebro, en la etiología de la esquizofrenia (Andreasen y Pearson, 2008; Shakiba, 2014). *LHX1*, por su parte, juega un papel importante en el control de la diferenciación de células de Purkinje en el cerebelo en desarrollo (Zhao y cols., 2007), así como en la diferenciación y función circadiana del núcleo supraquiasmático del hipotálamo (Bedont y cols., 2014). No obstante, su función ha sido estudiada más en profundidad en otros procesos de desarrollo tales como la nefrogénesis. Durante este proceso, *HNFI1B* regula directamente la expresión de *LHX1* (Drews y cols., 2011;

Lokmane y cols., 2010), lo que implica la existencia de una fuerte relación entre ambos genes. Por lo tanto, un mecanismo común podría estar detrás del riesgo conferido por las nueve SNV raras con potencial regulador en los pacientes con esquizofrenia que alteran las funciones de represor Polycomb en ambos genes, por lo que la implicación de las modificaciones epigenéticas en la esquizofrenia y su relación con la secuencia de ADN debe ser entendida para poder discutir estos resultados (Feinberg, 2007).

Son numerosas las evidencias que señalan a las modificaciones epigenéticas características de los distintos estados funcionales de la cromatina como factores de predisposición para el desarrollo de la enfermedad mental (Gavin y Sharma, 2010; Svrakic y cols., 2013). Particularmente numerosos son los estudios basados en la comparación de patrones de metilación de ADN en casos y controles sanos. Entre ellos, se incluyen la hipermetilación de los genes de las vías de GABA resultantes en neuronas corticales deficitarias en GABA en pacientes de esquizofrenia (Veldic y cols., 2004) o las diferencias de metilación de ADN en genes relevantes de importancia neurobiológica entre gemelos monocigóticos discordantes en esquizofrenia (Dempster y cols., 2011). Igualmente, Mill y cols reportaron patrones aberrantes de metilación en la secuencia de aproximadamente 100 *loci* asociados a la esquizofrenia, sugiriendo una disfunción epigenética sistémica en lugar de casos aislados (Mill y cols., 2008). Por otra parte, un estudio relevante en esquizofrenia demostró que los genes con diferencias significativas en los patrones de metilación de ADN en humano respecto a chimpancé mostraban una clara asociación con las categorías funcionales propias de enfermedades características del linaje humano como "trastornos neurológicos", "trastornos psicológicos" o "cáncer" (Zeng y cols., 2012).

Por otro lado, las modificaciones de histonas también han sido descritas en relación a la esquizofrenia, tal y como se demostró en varios estudios que encontraron niveles elevados de la enzima deacetilasa de histonas *HDAC1* en muestras *post mortem* del cerebro en esquizofrenia (Benes y cols., 2008; Sharma y cols., 2008). En esta línea, cerebros *post mortem* de mujeres esquizofrénicas presentaron niveles reducidos de metilación activadora en H3K4 en el promotor de *GAD67*, que codifica una enzima clave para la síntesis de GABA que está regulada por la actividad neuronal (Benson y cols., 1994), con una concordante reducción de la expresión génica (Huang y cols., 2007). Las metilaciones H3K4me3 y H3K27me3, activadoras y represoras de la transcripción, respectivamente (Mikkelsen y cols., 2007), y estudiadas a fondo en el proyecto ENCODE (Dong y cols., 2012; Sanyal y cols., 2012; Thurman y cols., 2012), también han demostrado una importancia específica en la regulación de la transcripción durante el desarrollo cerebral (Cheung y cols., 2010; Mikkelsen y cols., 2007).

La comparación sistemática de los niveles de las señales de H3K4me3 y H3K27me3, en las regiones promotoras (seleccionadas como 2000 pb alrededor del TSS) entre pacientes esquizofrénicos y controles sanos, identificó 22 genes cuya diferencia de expresión se correlaciona fuertemente con diferencias significativas de las señales H3K4Me3 y H3K27me3 entre ambos grupos (Kano y cols., 2013). No obstante, los estudios relativos a modificaciones histónicas como factores de riesgo a enfermedades complejas son más recientes y todavía menores en número a los de metilación de ADN, dada la mayor sencillez metodológica de estos últimos.

A pesar de las numerosas evidencias de la relación entre la alteración de los estados epigenéticos y la vulnerabilidad a enfermedades humanas, pocos estudios han centrado su investigación, hasta la fecha, en la influencia que ejerce la propia secuencia genética en la existencia de estos patrones epigenéticos modificados que condicionan la expresión y activación o silenciamiento de diversos genes. Esto es tremendamente importante, teniendo en cuenta que el estudio de las modificaciones epigenéticas entendidas como el mecanismo de interacción del ambiente con la cromatina, en lugar de como un producto de alteraciones en la propia secuencia de ADN, serviría para tratar de explicar la varianza asociada precisamente al ambiente. Precisamente, algunos trabajos recientes han sugerido que una gran parte de esas modificaciones histónicas u otras marcas epigenéticas que regulan los niveles de expresión tienen un origen eminentemente genético (Kilpinen y cols., 2013; McVicker y cols., 2013), y que, por tanto, es de enorme importancia el hecho de estudiar, junto con el epigenoma y los niveles de expresión, la variación genética que puede subyacer a estas alteraciones. Prueba de ello es que las variantes genéticas que afectan a elementos de regulación distales se relacionan tanto con cambios en la expresión como con modificaciones de la cromatina en las regiones promotoras de los genes que regulan (McVicker y cols., 2013).

Uno de los mecanismos epigenéticos bien conocidos, que regulan la expresión génica en eucariotas sin la implicación directa de la secuencia de ADN, es la adición de grupos metilo (CH₃) en la posición 5 de las citosinas, cuando éstas se encuentran unida a guanina formando dinucleótidos CpG, lo que provoca un silenciamiento de la expresión mediante la prevención de la necesaria unión de factores de transcripción y otras proteínas de unión al ADN, así como por el reclutamiento de proteínas inhibidoras del proceso de transcripción. A pesar de que este mecanismo se produce de forma general a lo largo de todo el genoma, hay tramos cortos conocidos como islas CpG, que generalmente se encuentran dentro o cerca de los sitios de inicio de transcripción y que presentan una alta proporción de dinucleótidos CpG no metilados, lo que favorece la expresión al dejar el ADN fácilmente accesible y disponible para la unión de los factores de transcripción necesarios (Blackledge y Klose, 2011). Dos tercios de las regiones promotoras se caracterizan por la presencia de estas islas CpG (Illingworth y Bird, 2009). Esta

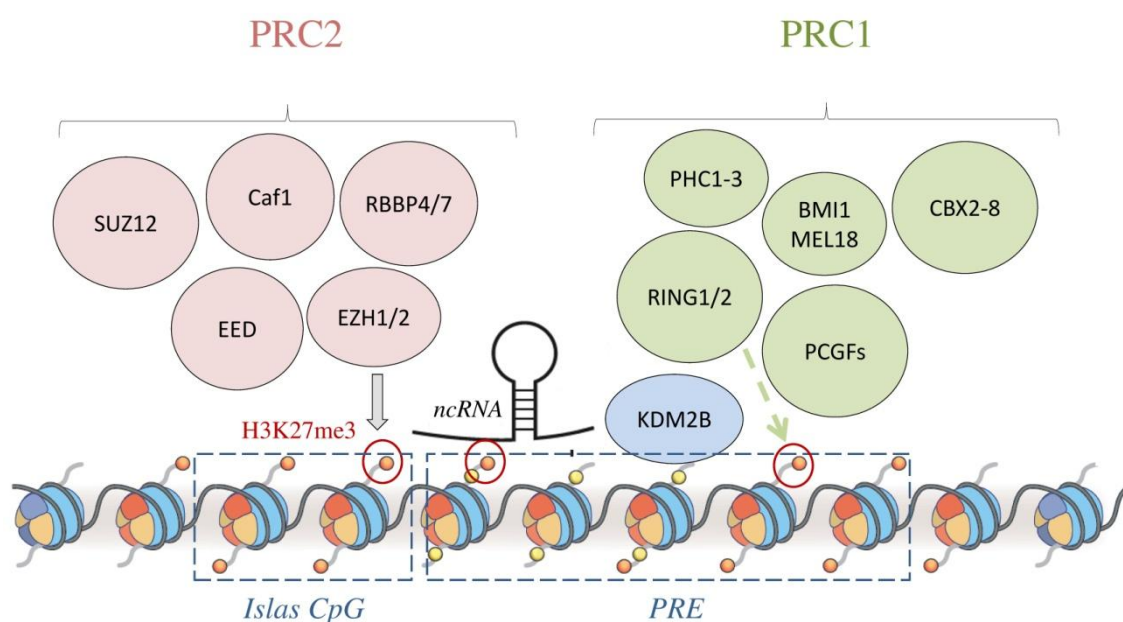
visión "pasiva" de las islas CpG ha sido recientemente cuestionada y se ha demostrado que tienen un papel más activo en todo este proceso, siendo capaces de reclutar proteínas que promueven la generación de ambientes más favorables para la iniciación de la expresión génica (Blackledge y cols., 2010; Thomson y cols., 2010). Uno de los grupos de proteínas que las islas CpG son capaces de reclutar son las que constituyen los denominados complejos represores Polycomb. Este grupo de proteínas fueron descritas inicialmente en base al fenotipo "polycomb", a raíz de un ejemplar mutante de *Drosophila melanogaster* que presentaba patrones anormales de segmentación (Lewis, 1949). Por extensión, aquellos genes cuyas disrupciones provocaban este fenotipo fueron considerados del grupo de represores Polycomb (Margueron y Reimberg, 2011). Este grupo de proteínas son, por tanto, necesarias para la regulación negativa de estos genes homeóticos (Lewis, 1978), es decir, aquellos que participan en el desarrollo de los organismos y que determinan la identidad de las partes individuales del embrión, imprescindible para la correcta segmentación del organismo, así como el mantenimiento de la expresión controlada de estos genes durante la edad adulta (Schuettengruber y Cavalli, 2009).

En mamíferos, se han descrito dos grupos de proteínas del complejo represor Polycomb, PRC1 y PRC2, que habitualmente actúan conjuntamente para reprimir la transcripción posicionándose sobre secuencias específicas en los promotores (conocidas como PREs, del inglés *Polycomb Response Elements*). Sin embargo, aunque el mecanismo está claramente identificado en *Drosophila*, en mamíferos no hay una evidencia tan clara. De hecho, ni PRC1 ni PRC2 presentan componentes con capacidad de unión al ADN. A diferencia de la mayoría de los factores de transcripción, una clave del funcionamiento de los complejos Polycomb es que, a pesar de que están presentes ubicuamente en todas las células, su unión a un gen diana específico depende de terceros elementos (**figura 5.6**).

El mecanismo por el cual se reclutan ambos PRC no se conoce al completo y las evidencias al respecto son todavía muy controvertidas (Schwartz y Pirrota, 2013). No obstante, se sabe que las islas CpG tienen capacidad para el reclutamiento de PRC2 (Lynch y cols., 2012; Mendenhall y cols., 2010), y que las proteínas EZH1 o EZH2 (esta última en mayor medida) son las responsables de la metilación H3K27me3 (Margueron y cols., 2008; Simon y Kingston, 2009). Esta modificación genera una compactación de la cromatina en esa región que imposibilita la transcripción (Grau y cols., 2011), al prevenir la acetilación en la misma lisina (H3K27ac), cuyo efecto es precisamente el contrario. Otro mecanismo propuesto para esta compactación es el que implica la unión preferencial de los integrantes de PRC1 a histonas con la modificación H3K27me3, por lo que PRC2 también contribuiría al reclutamiento de PRC1 (Ku y cols., 2008; Margueron y Reimberg, 2011). Sin embargo, también hay evidencias de su actuación de forma independiente (Schwartz y cols., 2006). Por otra parte, otro factor diferente, KDM2B, que se

une a regiones PRE (Farcas y cols., 2012), se ha visto implicado en el reclutamiento del complejo PRC1, demostrándose un mecanismo de actuación independiente de PCR2 (Farcas y cols., 2012). Otras proteínas de unión a ADN como YY1 o RIBP también se han visto implicadas en el reclutamiento del represor polycomb (Wilkinson y cols., 2006), que interaccionan con RING1 y RING2, los mismos factores implicados en la unión a la modificación H3K27me3. Finalmente, también hay varios indicios de que este complejo proceso de compactación esté favorecido en presencia de ARN no codificantes (Gupta y cols., 2010; Tsai y cols., 2010), con capacidad para reconocer secuencias específicas y reclutar elementos de PRC1 y PRC2. Una reciente revisión señala, en esta dirección, la relevancia de todas las especies no codificantes de ARN en la regulación de los procesos epigenéticos y su relación con la cognición y las patologías neuropsiquiátricas (Butler y cols., 2016). Otro estudio reciente también demostró la subexpresión de un ARN no codificante, Gomafu, en corteza cerebral de ratones sometidos a estímulos ansiógenos, lo cual tendría implicaciones en los mecanismos de represión Polycomb por la interacción de *Gomafu* con PRC1.

Figura 5. 6. Elementos que integran los complejos represores Polycomb y otros que intervienen en su reclutamiento y favorecen la compactación de la cromatina.



En cualquier caso, todavía se desconoce gran parte del mecanismo de actuación de todos los elementos que intervienen en la represión del Polycomb. Lo que sí se conoce en mayor medida es la dinámica de estos estados represores Polycomb a lo largo de la evolución. Existen numerosas evidencias acerca de los cambios que se producen en las marcas histónicas en todo el

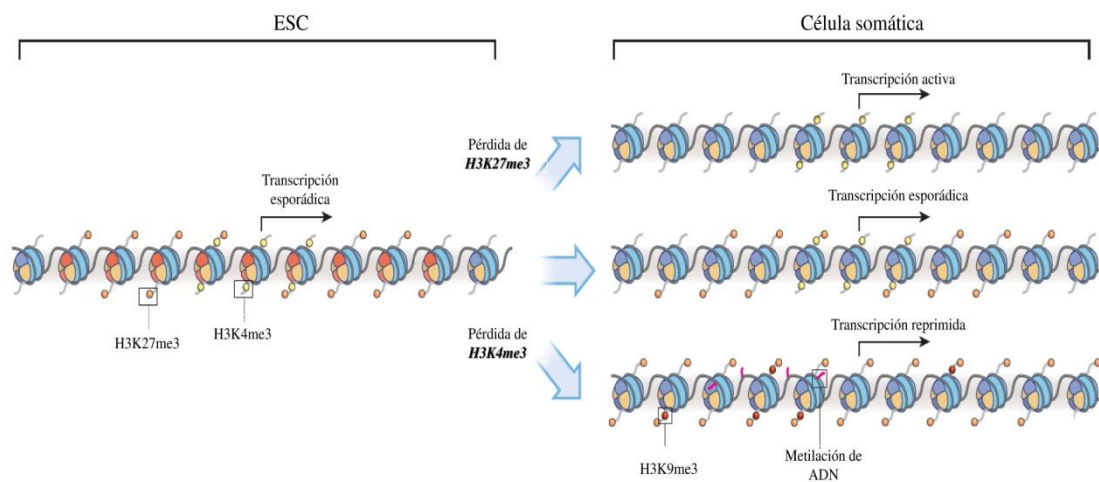
genoma y, por tanto, del dinamismo de los estados cromatínicos, consecuentes con los cambios transcripcionales producidos durante el desarrollo (Reik, 2007).

Muchos genes reguladores del desarrollo en las células madre embrionarias se caracterizan por la trimetilación de la lisina 4 de la histona H3 (H3K4me3), típica de las regiones transcritas, y, a su vez, por la trimetilación de la lisina 27 de la histona H3 (H3K27me3), típica de las regiones reprimidas. La combinación de estas marcas histónicas de funcionalidades opuestas en los promotores de genes implicados en el desarrollo les confiere a los mismos un bivalencia para actuar frente a estímulos variados, así como una plasticidad necesaria para la diferenciación de las células madre embrionarias (Margueron y Reimberg, 2011). Según la clasificación de los estados de la cromatina referidos en la base de datos del proyecto Roadmap (**tabla 1.9**), los promotores bivalentes son equivalentes a los estados “TSS Bivalente” (TssBiv) y “Flanco de TSS/*enhancer* bivalente” (BivFlnk), que difieren únicamente en la intensidad observada en otra metilación, en este caso de la lisina 4 de la histona H3 (H3K4me1), característica de regiones *enhancers* (Kundaje y cols., 2015). Durante el proceso de diferenciación de las células madre en cada uno de los linajes específicos, algunos *loci* pierden las marcas represivas H3K27me3, lo que conlleva la activación de la transcripción provocada por una mayor apertura de la estructura de la cromatina. Otros *loci*, por el contrario, pierden la metilación H3K4me3, consolidando un estado represivo más estable mediante la mayor compactación de la cromatina, y descrito en la base de datos del Roadmap como los estados “*Enhancer* bivalente” (EnhBiv) y “Represor Polycomb” (ReprPC) (**figura 5.7**). La marca de *enhancers* H3K4me1, presente y ausente en el estado EnhBiv y ReprPC, respectivamente, si bien no altera la funcionalidad eminentemente represiva del estado bivalente, permite la rápida transición a otro estado activo bajo la presencia de determinados estímulos externos (Kundaje y cols., 2015; Margueron y Reimberg, 2011).

En el estudio aquí discutido, las posiciones con variantes reguladoras en los genes *HNF1B* y *LHX1* de 17q12, evolucionan hacia una represión mayor en tejido adulto, tal y como se describe a partir de los datos de Roadmap para la evolución de los estados de la cromatina desde las células madre embrionarias (**figura 5.8**). Por lo tanto, el proceso de diferenciación temprana que tiene lugar durante el desarrollo del sistema nervioso central y del cerebro podría verse afectado en los nueve portadores de las variantes reguladoras (RegulomeDB “2”) mediante la interrupción de este proceso de represión, y aumentando, de este modo, el riesgo asociado a esquizofrenia. Esta hipótesis que se origina a la luz de los resultados obtenidos, es perfectamente coherente con el hecho de que la CNV 17p12 contribuye fuertemente a la susceptibilidad a diferentes trastornos del desarrollo neurológico, como la discapacidad intelectual o autismo, tal y como se comentó anteriormente (Malhotra y Sebat, 2012). Según esta premisa, el riesgo asociado a estos genes vendría de las alteraciones en sus niveles de expresión, siendo, probablemente, la delección más propia de trastornos severos, mientras que

una inhibición de la represión en estados post-embrionarios (y con ello, una expresión indebida en estadios en los que debiera estar reprimida) asociada al riesgo a trastornos más comunes y propios de la edad adulta como la esquizofrenia. En esta línea, el primer estudio discutido en relación al TCF4 ya hacía referencia a la haploinsuficiencia como factor de riesgo a un síndrome severo como el PHS y a variantes reguladoras con posible efecto aumentador de la expresión como factor de riesgo a esquizofrenia.

Figura 5. 7. Evolución de los estados bivalentes en células embrionarias pluripotentes a estados activos o reprimidos en etapas posteriores del desarrollo. Figura adaptada de Margueron y Reimberg (2011).

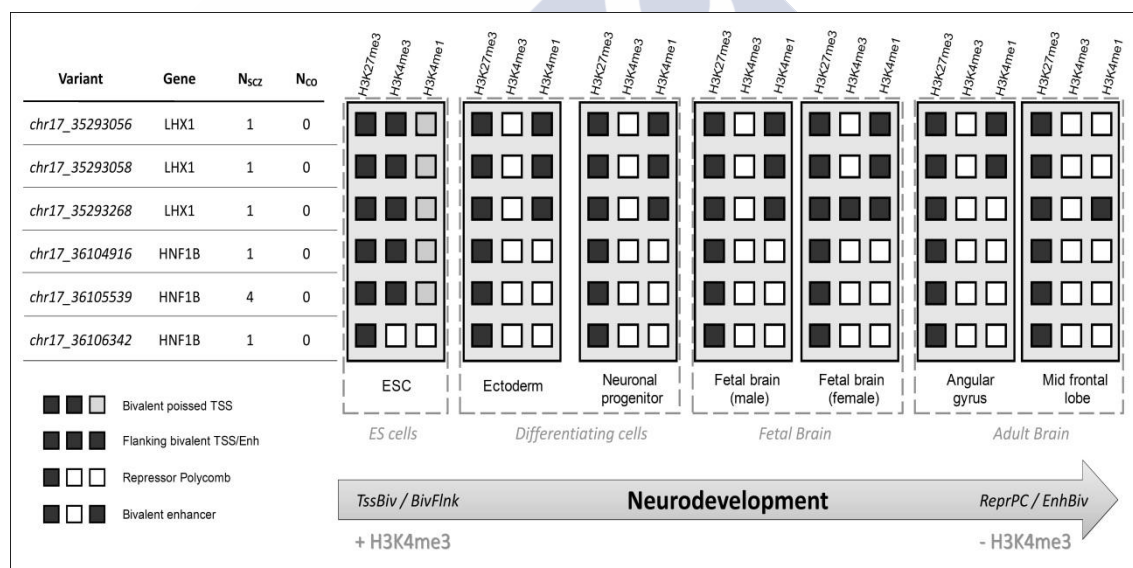


A pesar de lo novedoso de las hipótesis aquí presentadas en base a los resultados obtenidos, lo cierto es que recientes hallazgos en el contexto de la esquizofrenia apuntan a conclusiones parecidas. Hay suficientes evidencias como para poder afirmar que la remodelación de la cromatina mediante mecanismos de modificación de histonas juega un papel importante en el riesgo de esquizofrenia, y la genética representa un papel muy relevante a tenor de los datos aportados mediante diferentes fuentes como el estudio de mutaciones *de novo*, análisis de *pathways* o estudios de GWAS en esquizofrenia.

En este sentido, un estudio reciente reveló la presencia de dos variantes *de novo* de pérdida de función (LoF) en un gen con función de metiltransferasa en histonas, *SETDB1*, en pacientes con esquizofrenia mediante secuenciación de exoma de 231 tríos de enfermos (Takata y cols., 2014), un hecho altamente improbable por azar que, además, se suma a otras evidencias en esta dirección. Adicionalmente, la combinación de los datos del estudio de Takata y cols. con estudios previos de secuenciación de exoma en tríos de esquizofrenia reveló que la función de "regulación de la cromatina" era la más significativamente enriquecida en un análisis realizado sobre el conjunto de los 62 genes intolerantes a la variación y portadores de variantes *de novo*

patogénicas (Takata y cols., 2014). Curiosamente, entre los genes que pertenecen a este conjunto sobre los que se realizó el análisis de enriquecimiento hay dos implicados en la represión Polycomb, *BCORL1* y *KDM2B*, que presentan, además, una mutación *missense de novo* en un paciente con esquizofrenia en los estudios de Xu y cols. (Xu y cols., 2012a) y Girard y cols (Girard y cols., 2011), respectivamente. Ambos genes forman parte del complejo PRC1 que interacciona con PRC2 silenciando la transcripción. En el caso del *KDM2B*, además, es una de las piezas claves en el reclutamiento de los elementos de PRC1 a las secuencias PREs (figura 5.6) (Farcas y cols., 2012; Schwartz y Pirrota, 2013).

Figura 5. 8. Evolución de los estados de la cromatina durante el desarrollo basado en los datos del proyecto Roadmap Epigenomics. La representación de las líneas de células madre embrionarias (ESC) está basada en las marcas de modificación de histonas más comunes en las ocho líneas de ESC descritas en Roadmap (H1, H9, HUES6, HUES48, HUES64, ES-I3, ES-WA7 y UCSF-4star). Para el estudio del dinamismo de los estados de la cromatina durante el neurodesarrollo, se seleccionaron tres derivados de las células ESC diferentes de relevancia para el desarrollo del sistema nervioso central: células CD56+ de ectodermo derivadas de hESC, como una representación del ectodermo, y células de progenitor neuronal derivadas de ESC H1 y H9 como representantes de la etapa de progenitor neuronal. Ambas células progenitoras neuronales presentan las mismas marcas de histonas que se representan en la figura. Los tejidos de cerebro fetal (masculino y femenino) así como los tejidos adultos de giro cingulado y lóbulo frontal medio, se representan individualmente.



McCarthy y cols. llegaron a conclusiones similares tras el estudio de secuenciación de 57 tríos en esquizofrenia (McCarthy y cols, 2014), y en el que encontraron un exceso de mutaciones *de novo* LoF en genes implicados en la organización de la cromatina. Además, por otra parte, los autores extendieron estos hallazgos a los trastornos del neurodesarrollo relacionados, tales como trastornos del espectro autista o discapacidad intelectual, después de un análisis conjunto con datos de los estudios de secuenciación de exoma anteriormente realizados en tríos con esos trastornos.

Otra prueba que señala a la alteración de la función de remodelación de la cromatina como un factor de riesgo para los trastornos psiquiátricos procede del análisis conjunto de los resultados de estudios de GWAS relativos a esquizofrenia, trastorno bipolar y depresión mayor, mediante el que se identificaron las funciones "metilación de las histonas H3-K4" y "metilación de histonas", en general, como los *pathways* comunes más significativamente asociados al conjunto de estos trastornos en variación genética común (Network, T. & Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015). Información adicional acerca de la expresión génica condujo, igualmente, a la sugerencia de que la asociación a estos trastornos está impulsada principalmente por aquellos genes responsables de la metilación de histonas expresados principalmente durante la diferenciación neuronal y la determinación del tipo celular final. De nuevo, se llega a un escenario en el que un estudio conjunto utilizando diferentes fuentes de información vincula las diferencias de expresión, mediadas por variantes genéticas en regiones reguladoras no codificantes y que alteran los mecanismos epigenéticos, a la susceptibilidad a trastornos del neurodesarrollo como la esquizofrenia. De acuerdo con esto, Finucane y cols. desarrollaron un método para la partición de la heredabilidad basada en la contribución poligénica de las diferentes categorías funcionales de variantes en diferentes tipos celulares, teniendo en cuenta el efecto de desequilibrio de ligamiento en cada uno de los SNP implicados. La aplicación de este método para el mayor GWAS de esquizofrenia hasta la fecha (Finucane y cols., 2015) demostró un fuerte enriquecimiento de la proporción de la heredabilidad asociada a la esquizofrenia que se explicaba por aquellos SNP situados en los picos de ChIP-seq para la modificación histónica H3K4me3 en cerebro fetal a partir de los datos de ENCODE.

Finalmente, los *loci* más asociados a esquizofrenia en estudios GWAS también mostraron una implicación de los represores Polycomb en la etiología de la enfermedad. *EZH2*, que codifica para un metiltransferasa de histonas en el complejo PRC2, y que es la responsable de la trimetilación de la lisina 27 de la histona H3, H3K27me3 (Simon y Kingston, 2009), es una de las dianas experimentalmente validadas del gen asociado a esquizofrenia en los estudios más representativos *MIR137* (Ripke y cols., 2013a; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), tal y como se observó en la neurogénesis en ratón durante la edad adulta (Szulwach y cols., 2010). Este hecho es consistente con el incremento en el nivel de *EZH2* en líneas celulares linfoblastoides de pacientes con discapacidad intelectual asociado a la microdelección en 1p21.3, que afecta a los genes *DPYD* y *MIR137* (Willemsen y cols., 2011), ambos implicados en la etiología de la esquizofrenia. Precisamente, *EZH2* está presente en las señales de ChIP-seq para las posiciones de 8 de las 9 variantes de nuestro estudio (**tabla 4.7**). A pesar de que este factor no tiene capacidad reconocida de unión a ADN, su presencia asociada a las posiciones de estas variantes sugiere una gran importancia de las mismas en la metilación

H3K27me3 propia de represor Polycomb que caracteriza a estas regiones promotoras. De este modo, dada las recientes evidencias discutidas, es altamente probable que el mecanismo de reclutamiento de elementos de PRC1 y PRC2, y la consecuente metilación H3K27me3 que favorece la compactación de la cromatina, se vean afectados por estas mutaciones en posiciones de alta importancia reguladora (evidencias experimentales de una alta funcionalidad a raíz de estudio ChIP-seq, entre otras cosas, que llevan a una categoría RegulomeDB "2").

De hecho, varios estudios recientes han consolidado la influencia que presenta la alteración de los mecanismos de represión Polycomb en esquizofrenia o enfermedades relacionadas. Precisamente el gen *EZH2* ha demostrado estar directamente asociado a autismo en población China en la variante *rs6464926* (Li y cols., 2016a), que, además, es un eQTL descrito en la base de datos de Braineac (<http://www.braineac.org>) que influye en la expresión del gen en varias regiones cerebrales. No obstante, es necesario señalar que los resultados del trabajo de Li y cols. son preliminares, teniendo en cuenta el reducido tamaño muestral empleado. Otro estudio reveló una subexpresión en esquizofrenia de los microRNA de la familia *MIR320*, que tienen como una de las dianas validadas a *EZH2* (Vachev y cols., 2016). La subexpresión de este gen, de hecho, se ha visto implicada en la migración neuronal anormal debida a la consecuente sobreexpresión de *relina* (un gen cuya alteración de la expresión se asocia a desórdenes en el neurodesarrollo (Franco y cols., 2011)) de forma ectópica que provoca la interrupción de los mecanismos de represión Polycomb sobre este gen en ausencia de *EZH2* (Zhao y cols., 2015).

Es interesante señalar que la implicación del represor Polycomb en la etiología de la esquizofrenia, a tenor de los estudios discutidos así como del presente trabajo, permite hipotetizar acerca los indicios de epidemiologías contrapuestas de esquizofrenia y cáncer (Grinshpoon y cols., 2005). En este sentido, *EZH2* se emplea como marcador de la agresividad de los tumores de próstata y pecho, por la comprobada relación de su sobreexpresión con la progresión tumoral (Kleer y cols., 2003; Varambally y cols., 2002). De este modo, se podrían entender los resultados de los estudios que apuntan a una protección frente a cáncer de los pacientes esquizofrénica (Eskandari y cols., 2015; Lin y cols., 2013a). Son especialmente interesantes, desde el punto de vista de la genética, aquellos que encuentran esa protección en familiares de enfermos psiquiátricos (Ji y cols., 2013). En cualquier caso, dada la elevada poligenicidad de estos trastornos, esta hipótesis es todavía una mera especulación.

La principal limitación del presente trabajo es el hecho de no haber podido estudiar sistemáticamente todos los genes en las regiones de CNV implicados en el riesgo de esquizofrenia, así como la totalidad de genes relativos a *loci* asociados a la enfermedad en los estudios GWAS. Por otra parte, también hay que señalar, tal y como se describe en el apartado metodológico, que no se pudieron cubrir completamente todas las regiones promotoras relativas

a los TSS de los diferentes transcritos de cada gen seleccionado. A pesar de estos límites impuestos por razones eminentemente económicas, se ha demostrado que la estrategia de integración de los datos generados por dos grandes proyectos de reciente publicación, ENCODE y Roadmap, en el análisis de la variación no codificante, es altamente exitosa. Ésta es demostrable en tanto que las asociaciones encontradas revelan un riesgo superior al atribuible a las regiones codificantes, a pesar de que esos estudios tuvieran una potencia estadística francamente superior (Need y cols., 2012; Purcell y cols., 2014). Nuestra estrategia aquí discutida ha llevado, además, a la identificación de la remodelación de la cromatina, específicamente en lo que respecta a los estados represores Polycomb en 17q12, como un mecanismo implicado en la etiología de la esquizofrenia, y reivindica un mayor número de estudios acerca de los elementos de regulación represores Polycomb como factores de riesgo en el origen de los trastornos psiquiátricos y del neurodesarrollo. Si se validan con muestras adicionales, los hallazgos destacados en nuestro estudio podrían ser considerados como una prueba de concepto para el análisis de la variación no codificante en relación con la susceptibilidad a esquizofrenia, y especialmente relevante teniendo en cuenta la transición prevista de los estudios de secuenciación de regiones codificantes a aquellos que engloben la totalidad del genoma en un futuro próximo.

5.4. ESTUDIO IV. ANÁLISIS DE TRANSCRIPTOMA DE GIRO CINGULADO EN ESQUIZOFRENIA. ESTUDIO DE VARIACIÓN GENÉTICA Y DIFERENCIAS DE EXPRESIÓN EN CEREBRO

En el cuarto y último de los trabajos aquí presentados, se secuenció el transcriptoma de giro cingulado de 31 enfermos de esquizofrenia y 26 controles para realizar un análisis de expresión diferencial entre ambos grupos y un estudio de la variación genética en relación a los resultados del primero. En el estudio, realizado en gran parte en la *Columbia University* (Nueva York), se emplearon muestras bien caracterizadas del *Stanley Research Medical Institute* (SMRI), una institución sin ánimo de lucro que ha servido de muestras de gran calidad a estudios relevantes en esquizofrenia y otros trastornos psiquiátricos (Karlsson y cols., 2001; Sekar y cols., 2016; Tkachev y cols., 2003; Torrey y cols., 2005). En éste, se reveló una sobreexpresión de genes de astrocitos y subexpresión de genes de neuronas en esquizofrenia (figura 4.14), la presencia de variantes genéticas en casos que afectan a genes de neuronas específicos del neurodesarrollo (figura 4.29) y finalmente, un enriquecimiento de variantes en genes de astrocitos (tabla 4.15) en casos que explican parte de la variación de expresión observada en estos genes (tabla 4.18).

5.4.1. Discusión metodológica

Respecto al análisis de expresión, en primer lugar, es necesario señalar que el presente estudio de expresión se realizó en base a una hipótesis previa. Los perfiles de expresión en las regiones del cerebro humano se organizan de forma robusta en módulos de coexpresión que reflejan composición celular subyacente al tejido cerebral, lo que proporciona una nueva base para explorar las diferencias moleculares que distinguen a tipos celulares específicos y procesos funcionales implicados en las enfermedades de la mente humana (Oldham y cols., 2008). Dada la naturaleza de la esquizofrenia como enfermedad del neurodesarrollo (Najas-García y cols., 2014; Rapoport y cols., 2012), sumado a las evidencias que implican a los distintos tipos celulares (Bloomfield y cols., 2015; Catts y cols., 2014; Lewis y cols., 2005; Tkachev y cols., 2003) y módulos de genes coexpresados (de Jong y cols., 2012; Mistry y cols., 2013b; Ren y cols., 2015) en la etiología de esta enfermedad, en este trabajo se analizaron las diferencias de expresión en los transcriptomas de casos esquizofrénicos y controles sanos comparándolas con los módulos espaciotemporales de genes coexpresados en cerebro durante el neurodesarrollo descritos previamente (Hawrilycz y cols., 2012; Miller y cols., 2014), profundizando en la naturaleza específica de los tipos celulares más implicados con el análisis de los genes específicos de cada uno de ellos en nuestros datos de expresión (Cahoy y cols., 2008; Zhang y cols., 2014a).

Esta estrategia contrasta con la escogida por la mayoría de estudios relacionados, en los que se compara la expresión de los cientos de sets funcionales de genes definidos en bases de datos como KEGG (Kanehisa y Goto, 2000), BioCarta (Nishimura, 2001) o Gene Ontology (Ashburner y cols., 2000), sin ninguna hipótesis previa. A pesar de que esta estrategia agnóstica presenta la ventaja de considerar todas las posibilidades funcionales por igual, presenta muchos inconvenientes. Ya en estudios de GWAS, se sugirió la estrategia de análisis de datos en base a sets de genes preseleccionados, para evitar la corrección excesiva por múltiples pruebas intrínseca a las estrategias libres de hipótesis previas (Askland y cols., 2012), por lo que fue empleada con éxito en análisis de variantes en estudios posteriores (Costas y cols., 2013), e incluso en el estudio de exoma en esquizofrenia más importante hasta la fecha (Purcell y cols., 2014). Igualmente, en el ámbito de las enfermedades mentales, la utilización de *pathways* o sets comunes de genes se halla cuestionada porque la mayoría de los mismos fueron descritos en base a tejidos más disponibles, lo que podría conllevar una peor anotación de los sets en cerebro (Wang y cols., 2010b). Este hecho contrasta con el actual conocimiento de los patrones de coexpresión durante el desarrollo gracias a proyectos como BrainSpan (www.brainspan.org).

En segundo lugar, existen diferencias metodológicas importantes que se deben destacar en nuestro estudio de expresión respecto a la mayor parte de los realizados. Por un lado, el mapeo de lecturas frente al genoma de referencia se realizó con STAR (Dobin y cols., 2013), un *software* libre, de reciente desarrollo y de código abierto que es 15-20 veces más rápido que los demás programas más empleados, presenta menos falsos positivos, consigue unos valores de precisión equiparables a los del programa TopHat (Trapnell y cols., 2009) y detecta mayor número de uniones de *splicing* (Wang y cols., 2014). Probablemente por estas razones, STAR es el *software* recomendado por el *Broad Institute* en el análisis de datos de RNA-seq. Por otra parte, la diferencia de expresión de los genes entre casos y controles se evaluó basándose en recuentos de lecturas, en lugar de medidas normalizadas de RPKM o FPKM, ya que se ha demostrado recientemente que esta estrategia es más precisa cuando se buscan diferencias de expresión a nivel gen, en lugar de transcrito (**apartado 1.3.2.1**). Zhang y cols. compararon los programas DESeq, EdgeR y Cufflinks, siendo este último no recomendado para el análisis de expresión diferencial a nivel de gen, sobre todo si la profundidad de lectura es baja. Por el contrario, el empleo de EdgeR se demostró preferible para el análisis de la expresión diferencial a expensas de introducir potencialmente más falsos positivos, mientras que DESeq presentó los niveles más bajos de falsos positivos (Zhang y cols., 2014b). En el presente trabajo, para una mayor robustez de los resultados, sólo aquellos concordantes en ambos *software* fueron incluidos.

En el análisis de las diferencias de expresión entre casos y controles en los módulos de coexpresión y en los grupos de genes específicos de cada línea celular se emplearon dos

herramientas: GSEA (Subramanian y cols., 2005) y CSEA (Xu y cols., 2014). La primera fue la principal utilizada por su mayor potencia a la hora de analizar los datos, ya que utiliza la totalidad de los valores de expresión para buscar enriquecimiento en las listas de genes introducidas. De este modo, se utilizan todas las diferencias de expresión (y no solo aquellas significativas) para hallar la magnitud del enriquecimiento en los grupos de genes de los módulos de coexpresión o de líneas celulares. Por contrapartida, la herramienta de CSEA nos permite realizar un análisis del enriquecimiento espaciotemporal en el tipo celular y la región cerebral durante el desarrollo más detallado. Sin embargo, en ésta sólo se puede introducir listas de genes como *inputs* para hallar el enriquecimiento, por lo que, en nuestro estudio, esta herramienta fue empleada principalmente como confirmación de las observaciones halladas en GSEA mediante la utilización, exclusivamente, de los genes pertenecientes a los *loci* DE. Además, es necesario señalar que, a la hora del análisis de enriquecimiento en un tipo celular, no hubo datos de microglía, por lo que este tipo no pudo ser estudiado en profundidad como sí se hizo con astrocitos y oligodendrocitos.

Finalmente, nuestro estudio hizo uso de la exposición a medicación antipsicótica a lo largo de la vida, en base a los valores entregados por el *SMRI*, para discernir entre los genes correlacionados y aquellos que no se correlacionan con la medicación. Para ello, se realizó un estudio de correlación utilizando una metodología estadística muy potente como el *bootstrap* (Efron y Tibshirani, 1993) para el cálculo de intervalos de confianza en base a las correlaciones entre expresión génica y las dosis de mediación administradas. En este sentido, es importante señalar que la medicación antipsicótica es uno de los factores considerados de mayor interferencia en el análisis de expresión en esquizofrenia (Crespo-Facorro y cols., 2015; Martin y cols., 2015; Tan y cols., 2014), por lo que la implementación de una metodología que permita la selección de genes cuya variación en los niveles de expresión no se vea afectada por la medicación es de importancia mayúscula para el estudio de esta enfermedad. Sin embargo, lo inverso no tiene por qué ser cierto. Es decir, la correlación entre la medicación de por vida y la expresión génica podría ser de origen indirecto, y de este modo enmascarar los posibles cambios de expresión en determinados genes debidos a la condición de la enfermedad por el mero hecho de que los pacientes con dosis de los medicamentos más altas fuesen probablemente los fenotipos esquizofrénicos más graves. Bajo este escenario, los resultados obtenidos son conservadores, permitiendo desvelar cambios de expresión no causados por la medicación pero no descartar otros no visibles en los genes no correlacionados debido a las posibles asociaciones indirectas.

Otras variables de influencia demostrada en la expresión de numerosos genes en cerebro como el consumo de alcohol (Liu y cols., 2006a) o drogas de abuso (Torres y Horowitz, 1999), se hallaron desigualmente distribuidas en casos y controles en las muestras analizadas (**ANEXO**

I). Sin embargo, éstas han demostrado estar significativamente correlacionadas, a su vez, con la medicación antipsicótica, observándose dosis mayores de medicación en aquellos pacientes menos consumidores de alcohol y otras drogas, y viceversa (**figura 3.8**). De esta forma, su dependencia en los niveles de expresión quedaría englobada en la correlación con la medicación, permitiendo un análisis de los resultados de expresión libre de influencias externas, al referirse a aquellos no correlacionados.

5.4.2. Discusión del análisis de expresión

Mediante esta metodología descrita, un total de 1876 *loci* de 22386 con expresión detectable, definida como RPKM > 0,3 (Ramsköld y cols., 2009), se hallaron diferentemente expresados en esquizofrenia en las muestras de giro cingulado, estando 174 (9,2%) de ellos no correlacionados con la medicación. El análisis en GSEA para los módulos de expresión demostró una subexpresión y sobreexpresión de módulos enriquecidos en genes de neuronas y astrocitos (Hawrylycz y cols., 2012; Miller y cols., 2014), respectivamente, cuya significación se mantuvo incluso al eliminar los genes correlacionados con la medicación (**figura 4.14**), y se consolidó en análisis en sets de tipos celulares en estudios de distinta naturaleza metodológica (Cahoy y cols., 2008; Zhang y cols., 2014a). El análisis de CSEA reveló que los 95 *loci* DE sobreexpresados en esquizofrenia mostraron un fuerte enriquecimiento en líneas celulares de astrocitos corticales y un enriquecimiento en células gliales cerebelares (**figura 4.19**). Los 79 genes DE subexpresados, por el contrario, no revelaron los enriquecimientos esperados en líneas celulares neuronales (**figura 4.21**). No obstante, la no confirmación en CSEA de los resultados en GSEA no nos indican que no exista esa relación. Por una parte, la eliminación de genes correlacionados con la medicación podría conllevar la eliminación igualmente de aquellos correlacionados indirectamente. Por otra, tal y como hemos señalado anteriormente respecto a las limitaciones de esta técnica, sólo los genes de 174 *loci* DE no correlacionados con la medicación se utilizaron como *input* de entrada en comparación con la totalidad de los 2104 *loci* no correlacionados empleados por GSEA, y a diferencia de CSEA, en esta última sí se considera el orden en el enriquecimiento de los mismos. Los resultados de CSEA, de hecho, sí son significativos al incluir la totalidad de genes subexpresados (**figura 4.20**), ganando mayor potencia, y confirmando la tendencia observada en GSEA tanto para genes dependientes como independientes de la medicación administrada.

Esta asociación bimodal en genes de astrocitos y otras células gliales y genes de neuronas ya ha sido descrita anteriormente en autismo (Voineagu y cols., 2011; Xu y cols., 2014). En el trabajo de Voineagu y cols. se identificó, por un lado, un módulo de coexpresión neuronal subexpresado y enriquecido en los genes de susceptibilidad conocidos para autismo

como *FOXI*, y otro módulo sobreexpresado en cerebros de autismo enriquecido en genes de respuesta inmune y en marcadores de astrocitos y de microglia activada. Resultados similares fueron obtenidos, posteriormente, por Xu y cols., quienes además implementaron por primera vez la técnica de CSEA en datos de expresión de autismo (Xu y cols., 2014). Estos autores obtuvieron resultados parecidos a los aquí presentados, mostrando una subexpresión en genes neuronales en contraposición a la sobreexpresión en líneas celulares de astrocitos corticales e inmunológicas. Dada la similitud con estos resultados, nuestro estudio permite establecer un fuerte vínculo entre esquizofrenia y autismo en relación a los patrones de expresión observados en cerebro *post mortem*. Este solapamiento se suma a las numerosas evidencias que apuntan a la etiología genética común de estos trastornos (Carroll y Owen, 2009), sobre todo referidas a la haploinsuficiencia de genes contenidos en variantes estructurales como las CNV (Malhotra y cols., 2012; Rees y cols., 2013), y que fundamentan la noción de la esquizofrenia como trastorno del neurodesarrollo (Rapoport y cols., 2012). Es interesante señalar, además, que estos resultados bimodales integran en cierto modo las hipótesis neuroquímicas e inflamatorias propuestas en relación al origen de la esquizofrenia (**apartado 1.1.6**), aunque, obviamente, se trate de una simplificación que probablemente omita muchos factores a tener en cuenta.

En relación a los genes específicos de neuronas hallados subexpresados en esquizofrenia en este trabajo, son varios los estudios que han reportado con anterioridad hallazgos de alteraciones en la transmisión neuronal y en la arquitectura celular de la sinapsis en el contexto del neurodesarrollo en esquizofrenia (Johnson y cols., 2008; Hayashi-Takagi y Sawa, 2010). En este sentido, la esquizofrenia ya ha sido, desde hace años, caracterizada por una disfunción de la sinapsis en distintos tipos neuronales (Owen y cols., 2005b). Los estudios de expresión al respecto han demostrado, igualmente, una consistente subexpresión en esquizofrenia de los genes con importante actividad neuronal (Harrison y Weinberger, 2005; Hoftman y cols., 2015; Lewis y cols., 2005). El reciente estudio de Hoffman y cols. demostró, además, que las alteraciones de expresión de los genes del sistema gabaérgico en corteza cerebral de esquizofrenia no son consecuencia de los efectos acumulativos de la enfermedad, sino que es probable que se originen durante las etapas del desarrollo que preceden a la eclosión de la enfermedad (Hoftman y cols., 2015). En el estudio aquí discutido en giro cingulado en esquizofrenia, se replicaron las mismas tendencias: subexpresión de genes neuronales (**figura 4.14 - 4.15**) y aquellos implicados en procesos sinápticos (**figura 4.16**); pero, además, se demostró que ésta no es producto de la medicación antipsicótica a pesar de la influencia que los medicamentos pudieran ejercer sobre la expresión (Crespo-Facorro y cols., 2015; Martin y cols., 2015).

Respecto a los genes relacionados con células gliales (principalmente astrocitos) hallados sobreexpresados en el presente estudio, los resultados aportados en estudios anteriores

son bastante contradictorios en relación a la direccionalidad de las diferencias de expresión (Bernstein y cols., 2015). Mientras que la subexpresión de genes neuronales en esquizofrenia es un hecho más consolidado, en genes específicos de células gliales se ha descrito tanto subexpresión como sobreexpresión de los mismos (Bernstein y cols., 2015; Catts y cols., 2014; Katsel y cols., 2011). En los resultados aquí presentados, los genes de los módulos de coexpresión M10 (Hawrylycz y cols) y C38 (Miller y cols., 2014) se encontraron claramente regulados al alza, siendo ambos altamente enriquecidos en genes de astrocitos en las descripciones de sus respectivos trabajos. Este enriquecimiento se confirmó tanto en genes de líneas celulares cerebrales en GSEA como en el estudio de aquellos genes DE sobreexpresados mediante CSEA (**figura 4.19**). Los enriquecimientos observados reflejan, además, una distribución muy clara de los genes de astrocitos en el extremo de genes sobreexpresados (**ANEXO XI**). De hecho, el análisis de estos genes reveló un enriquecimiento en aquellos característicos de glía de Bergmann y oligodendrocitos de cerebelo, además de astrocitos de cortex y cerebelo, teniendo en cuenta los genes no correlacionados con la medicación. Sin embargo, en el análisis de tipos celulares en GSEA sólo se observaron las asociaciones con astrocitos, por lo que la evidencia con estos últimos es más consistente. Estos hallazgos, refuerzan los resultados de otros estudios anteriores que sugieren una implicación sistemática de estas líneas celulares en la etiología de la esquizofrenia mediada por anomalías en estos tipos celulares (Bernstein y cols., 2009; Bernstein y cols., 2015; Catts y cols., 2014; Katsel y cols., 2011; Patterson, 2009)

La células de la glía, en este sentido, están empezando a ser consideradas en el mismo nivel de importancia que las neuronas en el estudio de las enfermedades psiquiátricas, en tanto que recientes estudios han implicado a este tipo de células en varias funciones del cerebro que son altamente relevantes para los trastornos neuropsiquiátricos (Eroglu y Barres, 2010; Fields, 2008; Graeber, 2010; Sanacora y Banasr, 2013; Volterra y Meldolesi, 2005; Woodbury-Fariña y cols., 2014). A raíz de lo publicado actualmente, se ha postulado que astrocitos, microglia y oligodendrocitos pueden contribuir al desarrollo de trastornos psiquiátricos como la esquizofrenia o la depresión (Bernstein y cols., 2015; Patterson, 2009). Las principales evidencias al respecto provienen de los estudios de expresión, donde cada uno de los tipos celulares ha mostrado niveles alterados en esquizofrenia.

En particular, los oligodendrocitos han sido implicados desde hace años en la patofisiología de la esquizofrenia debido a que desempeñan el importante cometido de formar la vaina de mielina en el SNC. Alteraciones de la mielina se han implicado en el deterioro cognitivo y en el envejecimiento acelerado del cerebro propio de la esquizofrenia, así como en la plasticidad sináptica (Fields, 2005; Kochunov y cols., 2013). En un artículo reciente, se identificó un mecanismo a través del cual *Neuregulina (NRG)* y *BDNF*, ambos relacionados con

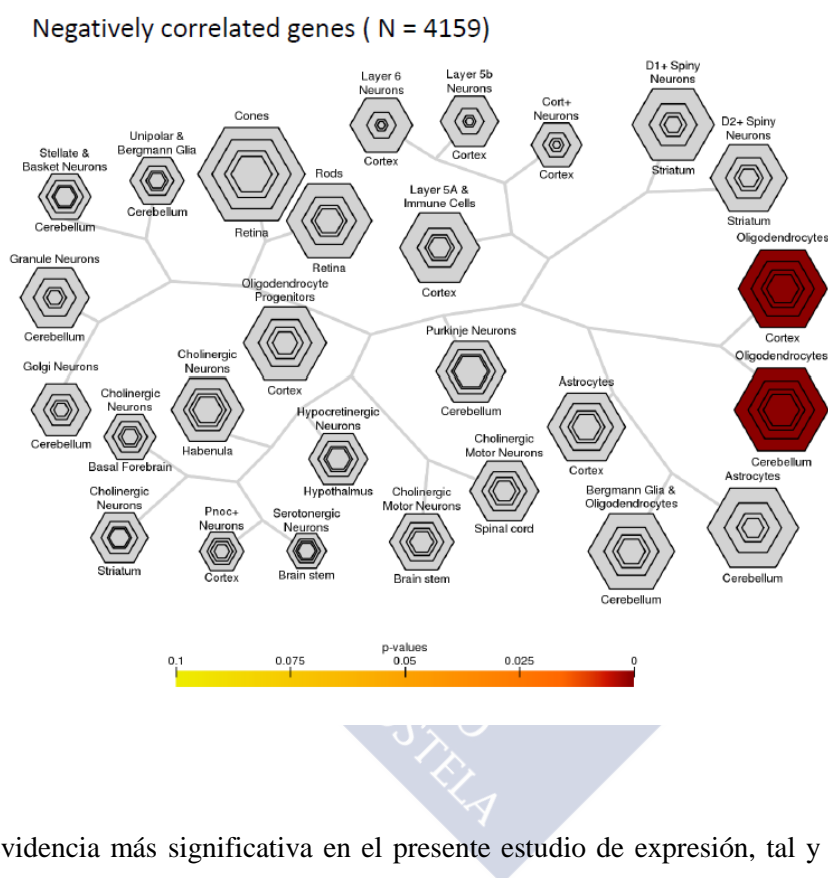
esquizofrenia, influyen en la mielinización al cambiar la actividad independiente de los oligodendrocitos por otra dependiente de la actividad neuronal de liberación de glutamato sobre los receptores NMDA (Lundgaard y cols., 2013). En varios estudios, se ha demostrado una reducción de la densidad de oligodendrocitos en la esquizofrenia que no se limita a la materia blanca cerebral, sino que también tiene lugar en la materia gris, incluyendo áreas corticales (Hof y cols., 2003) o del hipocampo (Schmitt y cols., 2009). Del mismo modo, numerosas evidencias sugieren que las alteraciones en el perfil de expresión de genes de oligodendrocitos observadas en esquizofrenia contribuyen a la etiología de la enfermedad (Goudriaan y cols., 2013; Bernstein y cols., 2015).

A pesar de que los resultados de CSEA sí apuntan a las líneas celulares de oligodendrocitos como unas de las enriquecidas en genes sobreexpresados en el presente trabajo, en GSEA no se observó esta tendencia ni en genes propios de líneas celulares de oligodendrocitos (Zhang y cols., 2014a) ni de módulos de coexpresión característicos de éstos (Hawrylycz y cols., 2012). Una de las principales razones de esta inconsistencia podría ser la gran influencia que ejerce la medicación sobre los genes de oligodendrocitos, como por ejemplo en el caso del Haloperidol, que contribuye a la reducción de la densidad de este tipo celular bloqueando la formación de oligodendrocitos maduros (Ren y cols., 2013). Es importante tener en cuenta que, del mismo modo que mediante el análisis de correlación entre medicación y expresión génica se pueden detectar anomalías probablemente debidas a la medicación, no se detectarían otros posibles efectos de la misma que, precisamente, corrigiesen anomalías propias de la enfermedad. En este sentido, todo parece indicar que la influencia de la medicación en nuestras muestras podría haber provocado este escenario en lo que se refiere a la expresión de genes de oligodendrocitos, tal y como sugiere el análisis en CSEA del total de genes expresados en giro cingulado e inversamente correlacionados con la medicación (**figura 5.10; ANEXO LV**), donde se alcanzan altos valores de significación como en oligodendrocitos de corteza cerebral. ($FDR-q = 2,78 \times 10^{-20}$ para pSI de 0,05). Además, en coherencia con las evidencias que sugieren siempre una reducción de la densidad de oligodendrocitos con la medicación (Ren y cols., 2013), ningún tipo celular se encuentra significativamente enriquecido cuando se analiza el total de genes directamente correlacionados con la medicación (**ANEXO LIII-LIV**). Por ello, es probable que la medicación contribuyera a la recuperación de los niveles normales de estos genes, equiparándolos a los de los controles sanos.

Las microglías, células neurogliales con capacidad fagocitaria que forman el sistema inmunitario del sistema nervioso central, ya se describieron anteriormente como asociadas a los procesos de neuroinflamación en varias regiones del cerebro en pacientes de esquizofrenia (Bloomfield y cols., 2015; Catts y cols., 2014; Takahashi y Sakurai, 2013), lo que ha conducido, incluso, a tratamientos anti-inflamatorios como coadyuvantes a la medicación clásica (Bernstein

y cols., 2015, Müller y cols., 2012). En nuestro estudio, el análisis de CSEA no incluyó este tipo celular, por lo que, sumado a la ausencia de enriquecimiento en GSEA (**figura 4.15**) a partir de los datos de líneas celulares (Zhang y cols., 2014a), no existen evidencias que sugieran la posible implicación de estas células en giro cingulado en esquizofrenia

Figura 5. 9. Enriquecimiento en tipo celular (CSEA) de los genes expresados en giro cingulado (RPKM > 0,3) e inversa y fuertemente correlacionados (FDR-p < 10⁻⁵) con la medicación. En el ANEXO LV se encuentran todos los valores de significación para cada uno de los casos analizados y en todas las regiones consideradas.



La evidencia más significativa en el presente estudio de expresión, tal y como hemos señalado, es la que implica una sobreexpresión de los genes de astrocitos en esquizofrenia, con independencia de la medicación y en todos los análisis realizados. Actualmente, se considera de enorme importancia el rol de los astrocitos en el funcionamiento del cerebro. Los astrocitos se aceptan actualmente como elementos cruciales en diversas funciones como el suministro de sustratos para el metabolismo de energía a neuronas y oligodendrocitos, el control del agua extracelular, la homeostasis de electrolitos, la expresión de neuromoduladores, la regulación de la liberación de neurotransmisores, la modulación de las respuestas inmunes, la producción de múltiples factores tróficos o la intervención para la migración de neuronas en desarrollo durante la regulación de la formación de sinapsis (Sofroniew y Vinters, 2010). Por ello, se considera que juegan un papel clave en la patogénesis de varios trastornos del desarrollo del SNC, incluyendo la esquizofrenia (Sofroniew y Vinters, 2010). Inicialmente, los estudios morfológicos reportaron

signos de astrogliosis en cerebro de esquizofrénicos (Bruton y cols., 1990; Stevens y cols., 1988), aunque posteriormente otros estudios no hallaron tales evidencias (Damadzic y cols., 2001; Falcai y cols., 1999), llegando en algunos casos a describir el escenario opuesto de reducción en el número de astrocitos (Williams y cols., 2013). De igual modo, en lo que respecta a estudio de expresión génica, existen hallazgos contrapuestos que dificultan el entendimiento de su rol en la etiología de la esquizofrenia. Frente a los estudios que revelaron una sobreexpresión de genes de astrocitos en esquizofrenia (Bruneau y cols., 2005; Catts y cols., 2014; Linderholm y cols., 2012; Matute y cols., 2005) y trastornos psiquiátricos relacionados (Voineagu y cols., 2011; Xu y cols., 2014), otros revelaron una regulación a la baja de los genes característicos de este tipo celular (Chandley y cols., 2013; Katsel y cols., 2011). Esta heterogeneidad en los resultados observados podría deberse al desigual comportamiento en las distintas áreas cerebrales (Garey, 2010) o a los efectos demostrados de la medicación sobre estas células gliales. En este último punto, un estudio en macacos demostró una significativa reducción del número de astrocitos causada por la exposición crónica a antipsicóticos (Konopaske y cols., 2008). Sin embargo, también se han reportado influencias en la dirección opuesta, como la reducción de la densidad astrocítica y de oligodendrocitos en ratas tratadas con antipsicóticos (Vernon y cols., 2014).

En el presente estudio, el análisis de CSEA de todos los genes expresados fuertemente correlacionados no reveló ninguna asociación (ni directa ni inversa), contrariamente a lo que ocurría con oligodendrocitos. Sin embargo, las líneas celulares de astrocitos sí aparecen significativamente enriquecidas en CSEA cuando se analizan genes sobreexpresados en esquizofrenia tanto inversa como directamente correlacionados con la medicación (**ANEXO LVII-LX**), revelando que la influencia de la medicación se restringe a un grupo menor de genes (aquellos DE) y tiene lugar de forma bidireccional, concordando con la divergencia en la literatura al respecto (Konopaske y cols., 2008; Vernon y cols., 2014). No obstante, también se mantiene una fuerte significación cuando se analizan los genes sobreexpresados independientes de la medicación (**figura 4.19**), por lo que, a raíz de los resultados del presente estudio, podemos afirmar que, si bien la medicación puede ejercer influencia sobre algunos genes de astrocitos, la implicación de este tipo celular en esquizofrenia va más allá de la consecuencia de los antipsicóticos administrados.

Como punto final a este bloque de discusión de los resultados de expresión, es importante destacar que entre los genes expresados en giro cingulado (RPKM > 0,3), existe un notable 37% de genes (8935 *loci*) no codificantes de proteínas (**tabla 4.8**), a pesar de las evidencias demostradas de que estos se expresan, en comparación con los genes codificantes, en unos niveles que solapan con el umbral de detección fijado en el estudio actual (Djebali y cols., 2012). Esto demuestra la importancia de trabajar con una base de datos como Gencode que, a

diferencia de otras como RefSeq, tiene en cuenta un gran número de transcritos novedales y un mayor número de genes no codificantes descubiertos durante el proyecto Encode, e incluso ha sido recientemente recomendada en la búsqueda de variantes funcionales (Frankish y cols., 2015).

A pesar de que el porcentaje de *loci* DE no codificantes es considerablemente menor (15,5%; **tabla 4.12**), un total de 291 *loci* DE presentan esta naturaleza, revelando la existencia de un gran número de especies DE en esquizofrenia con funciones eminentemente reguladoras, y en línea con la mayor importancia que han adquirido los genes de miRNAs (Beveridge y cols., 2010; Kocerha y cols., 2009) u otras especies no codificantes en el desarrollo de enfermedades mentales (Barry y cols., 2014; Liu y cols., 2014).

De los 15 genes más DE en este trabajo (**tabla 4.13**), en primer lugar se encuentra *NeuroD6*. Este gen codifica para un factor de transcripción de la familia estructural hélice – giro –hélice (HLH), de igual modo que los otros dos factores de transcripción estudiados en los dos trabajos iniciales presentados en este proyecto de tesis *TFC4* (**Estudio I**) y *NPAS3* (**Estudio II**). *NeuroD6* se encuentra implicado en el desarrollo y diferenciación del sistema nervioso, y fue hallado DE en estudios anteriores en corteza prefrontal en esquizofrenia (Pérez-Santiago y cols., 2012). De hecho, el análisis de su expresión en cerebro a lo largo del desarrollo revela, del mismo modo que en *TCF4* y *NPAS3*, que los máximos niveles de expresión se encuentran en la etapa del desarrollo prenatal, reforzando su papel como regulador de la transcripción en fases iniciales del neurodesarrollo humano. Además, también se ha demostrado su implicación en la tolerancia a estrés oxidativo al ser responsable de la activación de mecanismos de respuesta antioxidante (Uittenbogaard y cols., 2012), lo que permitiría hipotetizar sobre posibles vínculos entre su subexpresión y el aumento del estrés oxidativo característico de la esquizofrenia (Bitanhirwe y Woo, 2011). Por otro lado, este gen se ha demostrado como un marcador para la enfermedad de Alzheimer (Satoh y cols., 2014), lo que es coherente con la visión neurodegenerativa de la propia esquizofrenia que conlleva un deterioro cognitivo progresivo. Entre estos 15 genes, de hecho, *TAC1*, que codifica para cuatro productos hormonales que se cree funcionan como neurotransmisores, y *NDUFA4*, que codifica para una proteína implicada en la cadena respiratoria, también se hallaron DE en relación con la enfermedad de Alzheimer en estudios muy recientes (Li y cols., 2016b; Sekar y cols., 2015). También es destacable la presencia del gen de *GAD2*, subexpresado en este estudio y que codifica para la descarboxilasa del ácido glutámico, que controla la síntesis de GABA junto con *GAD1*. *GAD2* ya se halló, anteriormente, subexpresado en esquizofrenia (Davis y cols., 2016), por lo que este estudio refuerza la importancia del sistema gabaérgico en la etiología de la esquizofrenia. En cualquier caso, los 15 genes más DE representados en la **tabla 4.13** se encuentran, en su totalidad,

fuertemente correlacionados con la medicación, por lo que no se puede descartar que estas diferencias de expresión sean producto de la misma.

Un análisis de los genes más DE y no correlacionados con la medicación (**tabla 4.14**) también reveló resultados interesantes en relación a la esquizofrenia. El más DE y no correlacionado fue *NHP2L1*, una proteína nuclear que juega un papel importante en el mecanismo de *splicing* (Liu y cols., 2006b) y presenta metilación específica de alelo (Fang y cols., 2012), aunque su función no se encuentra bien definida tanto en la fisiología normal como en un proceso de enfermedad. En segundo lugar se encuentra el factor de transcripción implicado en el desarrollo del tejido nervioso *PAX6*, sobreexpresado en este trabajo, y que se expresa en poblaciones selectivas de las neuronas dopaminérgicas y ha demostrado ser un factor de protección frente a la pérdida de neuronas dopaminérgicas en la enfermedad de Parkinson (Thomas y cols., 2015), por lo que es coherente con la hipótesis dopaminérgica de la esquizofrenia. En tercer lugar, y probablemente conformando el resultado más importante, se encuentra *INA*, que codifica para uno de los neurofilamentos responsables de la formación del exoesqueleto de la neurona. Este gen, subexpresado en el actual estudio, forma parte del tercer *loci* de los 108 asociados a esquizofrenia (sólo por detrás de *miRNA137* y *DPYD*) en el estudio de GWAS más importante hasta la fecha (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) con un alto valor de significación ($P = 6,12 \times 10^{-19}$), así como en el estudio inmediatamente anterior, de menor potencia (Ripke y cols., 2013a). Por último, es interesante también la presencia del gen *CPLX2*, que codifica para una proteína implicada en la exocitosis de las vesículas sinápticas y se encuentra subexpresado en nuestro estudio. La regulación a la baja de este gen ya se fue reportado anteriormente en el contexto de la esquizofrenia (Knable y cols., 2004). Además, variantes genéticas en su secuencia se han asociado recientemente a la enfermedad en población coreana (Lee y cols., 2005), además de a la disfunción de la memoria (Hass y cols., 2015) y la pérdida cognitiva (Begemann y cols., 2010) propias de la enfermedad.

5.4.3. Discusión del análisis de variantes genéticas

El análisis de variantes raras (singletons de frecuencia $< 1\%$ en bases de datos disponibles) en los módulos de astrocitos y neuronas asociados en el estudio de expresión demostró, por un lado, un enriquecimiento en casos respecto a controles de variantes en genes de módulos subexpresados específicos de neuronas y regiones corticales a lo largo del neurodesarrollo en casos esquizofrénicos y, por otro lado, un enriquecimiento en casos respecto a controles de variantes en genes de módulos sobreexpresados específicos de astrocitos, estableciendo una clara conexión entre la variación genética y los resultados de expresión que

conforman dos evidencias: la implicación del desarrollo neuronal y de las células gliales (especialmente astrocitos en este estudio) en esquizofrenia. Es importante señalar que los genes DE subexpresados en esquizofrenia en este estudio presentan los valores más altos de expresión (**figura 4.17**), por lo que tiene sentido el enriquecimiento tan significativo de variantes que albergan los mismos (**tabla 4.15**) al seleccionar posiciones con $DP > 10$ y $RPKM > 0,3$, al producirse un sesgo metodológico debido a la inclusión de un mayor número de posiciones en los genes más expresados (estos son, los subexpresados en esquizofrenia en el estudio actual).

Respecto a la variación genética que afecta a los genes pertenecientes a módulos subexpresados, nuestros datos demostraron la disrupción de los genes neuronales durante el neurodesarrollo, reforzando resultados previos que describían mutaciones en genes coexpresados durante el desarrollo cortical desde etapas fetales a otras posteriores mediante el análisis de mutaciones *de novo* (Xu y cols., 2012a, Gulsuner y cols., 2013). En nuestro estudio, replicamos esas asociaciones analizando exclusivamente variación presente *in situ* en transcriptoma cerebral, por lo que esto sugiere que la búsqueda de variantes en genes seleccionados por sus patrones de expresión es una buena estrategia incluso en condiciones de baja potencia estadística. En este sentido, en esquizofrenia, también se ha descrito una acumulación de mutaciones *de novo* en genes relacionados y altamente expresados en corteza cerebral, particularmente durante el desarrollo prenatal (Xu y cols., 2011a; Gilman y cols., 2012). Igualmente, los genes mutados de módulos subexpresados en esquizofrenia se encontraron enriquecidos en varias etapas del desarrollo cortical pero, además, otro tanto fue observado en tejido talámico durante la etapa neonatal. A pesar de que el área cerebral más implicada en esquizofrenia ha sido el córtex prefrontal (Meyer-Lindenberg, 2010; Salgado-Pineda y cols., 2004), el posible rol del tálamo en esquizofrenia también ha sido explorado en diversos estudios, sobre todo mediante técnicas de neuroimagen (Pergola y cols., 2015; Schmitt y cols., 2004). Esta región tiene un papel crítico en el filtrado de la información sensorial y en la regulación de la entrada de estímulos cognitivos en la corteza cerebral. En esquizofrenia, hay evidencias consolidadas de que la disfunción del circuito cortico-talámico contribuye a la disimetría cognitiva característica de la enfermedad (Andreasen y cols., 1999), por lo que la disrupción de genes de este área cerebral en estadios iniciales del desarrollo tiene cabida en una concepción de la esquizofrenia como trastorno del neurodesarrollo. Sin embargo, a nuestro conocimiento, no existen evidencias actualmente que vinculen la variación genética directamente con el desarrollo de esta área cerebral en el ámbito de la esquizofrenia.

El análisis de CSEA de los genes mutados de módulos subexpresados nos ha permitido, además, revelar un enriquecimiento significativo de las mutaciones que afectan al tipo celular neuronal de proyección cortico-talámica Ntsr+ (**ANEXOXXXI-XXXXII**), algo que concuerda con el enriquecimiento en las áreas corticales y talámicas implicadas en el

neurodesarrollo. El tipo celular de neuronas Ntsr+ hace referencia a aquellas que contienen el receptor 1 del neuromodulador Neurotensina (NT), *NtsR1*, distribuido a lo largo de todo el SNC y con especial importancia en la transmisión dopaminérgica (Mai y cols., 1987), ya que *NT* actúa sobre su receptor reduciendo la función fisiológica de los receptores de dopamina (Jomphe y cols., 2006). En relación con esquizofrenia, estudios anteriores han sugerido un papel relevante para *NT* y *NtsR1* en la enfermedad. La densidad de los receptores de *NT* se encontró reducida en pacientes con esquizofrenia (Hamid y cols., 2002). Además, los niveles de *NT* en líquido cefalorraquídeo se hallaron inversamente relacionadas con la gravedad de la psicosis en esquizofrénicos no tratados, mientras que el aumento de los niveles de *NT* se asoció con una mejoría en los síntomas durante el tratamiento (Sharma y cols., 1997). Variantes genéticas en este receptor se han encontrado asociadas a una reducción de la memoria operativa (en inglés, *working memory*), un endofenotipo de la esquizofrenia (Li y cols., 2011; Li y cols., 2013). De hecho, un estudio posterior en población china demostró que tres polimorfismos previamente asociados con memoria operativa se encontraban, igualmente, asociados a esquizofrenia (Ma y cols., 2013). Curiosamente, este tipo celular es, también, uno de los que están enriquecidos en genes con variación genética descrita en estudios de autismo (Xu y cols., 2014).

Respecto a las variantes detectadas en genes pertenecientes a módulos sobreexpresados en esquizofrenia, en el estudio aquí presentado se observó una acumulación significativa de 23 variantes en los 29 casos respecto a las 3 halladas en 17 controles (FDR- $p = 0,032$, valor p del test de Cochran-Armitage = $4,6 \times 10^{-3}$) en los módulos de coexpresión C38 (Miller y cols., 2014) y M10 (Hawrylycz y cols., 2012), descritos en transcriptoma prenatal y durante el desarrollo hasta la adultez, respectivamente. Con estos resultados, se demuestra un vínculo entre genes de astrocitos regulados positivamente en esquizofrenia y la acumulación de variantes raras en casos respecto a controles (**tabla 4.13**).

Uno de los grandes dilemas respecto a las evidencias de la contribución de esta línea celular a la patofisiología de la esquizofrenia es, de hecho, si las perturbaciones descritas responden a un mecanismo secundario de inflamación consecuencia de la neurodegeneración o si contribuyen al riesgo de forma directa (Goudriaan y cols., 2013). En este sentido, un estudio reciente reveló que las anormalidades en los transportadores de glutamato de astrocitos conducen a un efecto compensatorio por el cual las neuronas sobreexpresan estos transportadores, y que, además, es independiente de la medicación (McCullumsmith y cols., 2015). En el ámbito de los estudios genéticos, un estudio reveló un fuerte enriquecimiento de los *loci* de GWAS en esquizofrenia (Ripke y cols., 2013a) en sets de genes específicos de astrocitos, constituyendo la primera evidencia de la disfunción genética de estas células como un factor de riesgo a la enfermedad (Goudriaan y cols., 2013). Estos autores, utilizando una muestra de 13689 casos y 18226 controles sanos (Ripke y cols., 2013a), demostraron que la variación en genes de

astrocitos ($P = 5,17 \times 10^{-19}$) y oligodendrocitos ($P = 3,43 \times 10^{-12}$) se hallaban fuertemente asociados a esquizofrenia, algo que no ocurría con genes de microglía ($P = 0,09$). Las asociaciones significativas se repartieron, además, en 6 y 3 sets de genes de astrocitos y oligodendrocitos, respectivamente, donde se observó la implicación genes específicos para cada tipo de célula, e incluso función específica, apoyando así los diferentes papeles que estos tipos celulares presentan en la patología de la esquizofrenia.

En el caso del módulo C38, se detectaron 8 variantes en casos y ninguna en controles (tabla 4.14), a lo largo de tres genes (5 mutaciones en *PCBD2*, una en *BCL2L14* y dos en *CAMTA1*). 4 de las 5 mutaciones de *PCBD2*, un gen de unión a proteínas que no ha sido reportado con anterioridad como factor de riesgo para enfermedades neuropsiquiátricas en nuestro conocimiento, son sitios diana de la proteína FMRP según evidencias *in vitro* (Ascano y cols., 2012), lo cual es muy interesante en el contexto de la esquizofrenia debido a las numerosas implicaciones de las mutaciones en las dianas de esta proteína en el riesgo a padecer esta enfermedad (Fromer y cols, 2014; Purcell y cols, 2014, Kirov y cols., 2012), así como en trastornos del espectro autista (de Rubeis y cols, 2014). El gen *PCBD2* se encuentra solapado con *CATSPER3*, por lo que se incluyeron ambos en el mismo locus de expresión. Este último gen codifica para un canal no selectivo de cationes de la familia *CATSPER* de expresión restringida a los testículos y esperma (Qi y cols., 2007). Sin embargo, con posterioridad al estudio de Qi y cols. se reveló que su expresión se extendía a varias regiones cerebrales en ratón, y que lo hacía de forma notoria en presencia de altos niveles de estrés oxidativo (Kim y cols., 2010). Por otra parte, *CAMTA1*, activador de la transcripción mediante la unión a calmodulina-1, que presentó dos deleciones en el presente estudio, ha sido asociado previamente con discapacidad intelectual (ID) y ataxia cerebelosa (Mikhail y cols., 2011. Thevenon y cols, 2012). La supresión de este gen, además, también ha sido asociada con el desarrollo de autismo (Pinto y cols., 2010). Una última variante tiene lugar en el gen *BCL2L14*, miembro de la familia de proteínas BCL2 que actúan como reguladores anti o pro apoptóticos y que están involucrados en una amplia variedad de actividades celulares. La sobreexpresión de este gen, además, se ha demostrado que induce la apoptosis en las células (Pickard y cols., 2011).

Aunque, por sí solo, la acumulación de variantes en el módulo C38 no alcance significación estadística, es destacable el hecho de que 8 variantes no codificantes estén presentes en casos y ninguna en controles, siendo además este módulo de coexpresión el más sobreexpresado en los datos de expresión y muy enriquecido en diversas funciones de regulación en los procesos de replicación y ensamblaje de cromatina durante el desarrollo prenatal del cerebro, estableciendo un vínculo importante con la hipótesis del neurodesarrollo y la regulación de la expresión génica en el contexto de la esquizofrenia. En concreto, en el análisis de las categorías funcionales de KEGG, PANTHER y *Gene Ontology*, destaca la

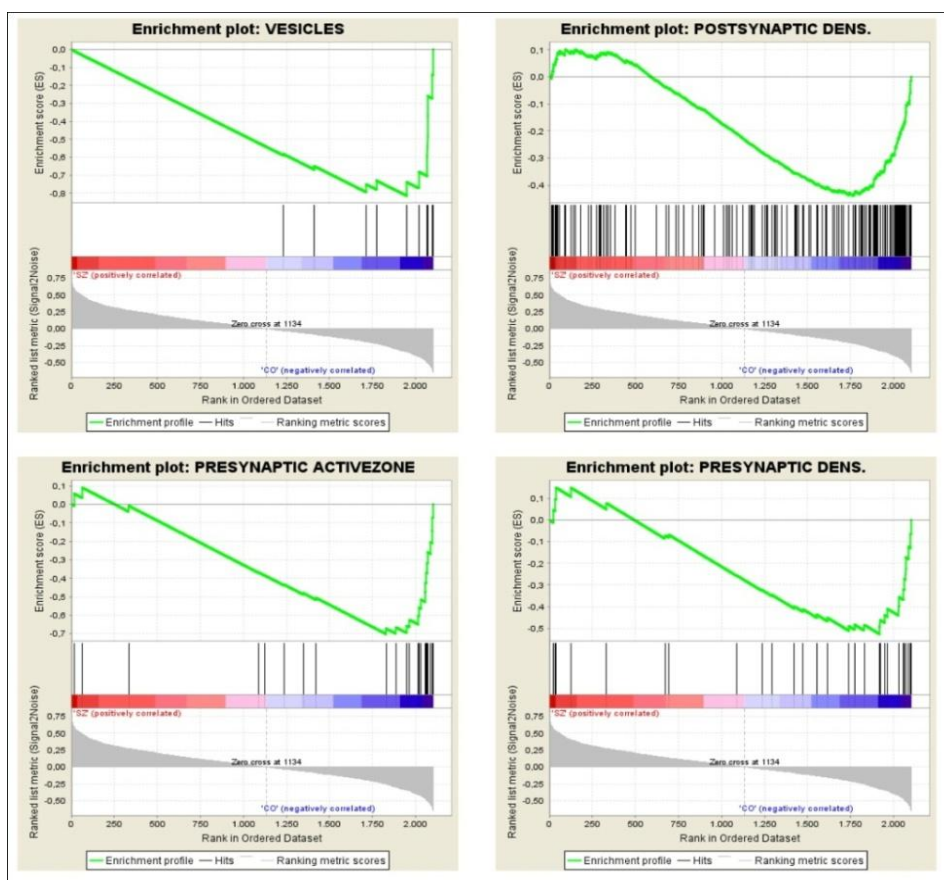
elevada significación (corregida) del enriquecimiento de este módulo en aquellos genes implicados en procesos de la división celular como la metafase ($P = 9,34 \times 10^{-63}$), mitosis ($1,85 \times 10^{-50}$), meiosis ($2,20 \times 10^{-13}$), la segregación de cromátidas hermanas ($1,38 \times 10^{-12}$), ensamblaje de cromatina ($7,59 \times 10^{-5}$), ensamblaje de centrómeros ($7,98 \times 10^{-4}$) o reparación de ADN ($5,71 \times 10^{-6}$) (Miller y cols., 2014).

Tanto los genes mutados como la funcionalidad del conjunto de aquellos que conforman el módulo C38 sugieren una gran importancia de los procesos básicos de regulación durante el neurodesarrollo prenatal. Por esta razón, estos resultados apuntan en la misma dirección que los descritos en el **estudio III**, anteriormente discutido, donde se demostraba la enorme importancia de las variantes genéticas que afectan a modificaciones cromatínicas en el contexto de la esquizofrenia y trastornos del neurodesarrollo (**apartado 5.3**). Además, son tremendamente coherentes con los del trabajo de Goudriaan y cols., donde varios conjuntos de genes de astrocitos asociados a esquizofrenia se componen principalmente de factores de transcripción y genes implicados en mecanismos de control epigenéticos (Goudriaan y cols., 2013). De hecho, casi la mitad de los genes descritos en ese estudio fueron previamente implicados en procesos de desarrollo. La gran contribución de los genes de control epigenéticos en los sets de genes transcritos en astrocitos (Cahoy y cols., 2008), donde se encuentran genes implicados directamente en esquizofrenia como *SETDB1* (Takata y cols., 2014), está en línea con la función propuesta de los mecanismos epigenéticos en esquizofrenia (Insel y cols., 2010). Los mecanismos de metilación del ADN y modificación de histonas en etapas iniciales del neurodesarrollo se han descrito, en este sentido, como "cicatrices" de experiencia temprana (Insel y cols., 2010), donde los factores genéticos modularían esa relación con el ambiente.

Es interesante señalar, igualmente, que el módulo C38 también se encuentra enriquecido en la categoría de genes de la densidad postsináptica ($P = 2,55 \times 10^{-4}$; Bayés y cols., 2011). Aunque en principio podría parecer sorprendente teniendo en cuenta la preconcebida naturaleza neuronal de la sinapsis, los últimos estudios al respecto han demostrado que precisamente los astrocitos contribuyen de forma significativa a la misma (Eroglu y Barres, 2010). Los procesos bioquímicos características de estas células se encuentran en gran medida asociados con sinapsis, y esta asociación es a la vez estructural y funcional. En el hipocampo, los astrocitos protoplasmáticos se encuentran recubriendo el 57% de las sinapsis, la mayoría de las cuales son las sinapsis excitatorias. Además, se ha demostrado que las sinapsis con mayor liberación de glutamato son más propensas a estar cubiertas por astrocitos (Witcher y cols., 2007). Sus transportadores recogen el glutamato liberado en la sinapsis y lo devuelven a las neuronas en forma de glutamina. En este sentido, la gliogénesis y la sinaptogénesis se producen simultáneamente en el cerebro, y maduración de las células gliales marca el final de los períodos sinaptogénicos (Fields, 2005), por lo que ambos procesos están íntimamente relacionados.

Neuronas cultivadas en presencia de una capa de astrocitos experimentaron un incremento del número de sinapsis de hasta 10 veces (Pfrieger y Barres, 1997). Igualmente, estas líneas celulares han demostrado ser necesarias para la inducción de la formación de sinapsis en neuronas generadas a partir de células madre embrionarias (Wu y cols., 2007).

Figura 5. 10. Enriquecimiento de genes implicados en la sinapsis de SynaptomeDB (Pirooznia y cols., 2012) en los datos de expresión de genes no correlacionados con la medicación. En cada una de las cuatro gráficas (de izquierda a derecha y de arriba a abajo: vesículas sinápticas, densidad postsináptica, zona activa presináptica y densidad presináptica), se representa la posición de los genes de cada conjunto respecto al orden del total de 22386 *loci* expresados, yendo de más sobreexpresados (rojo, izda.) a más subexpresado en esquizofrenia (azul, dcha.).



En nuestro estudio, los genes de sinapsis, recogidos de la base de datos SynaptomeDB han demostrado estar globalmente subexpresados, e independientemente de la medicación (tabla 4.16). Por otra parte, analizando las curvas de enriquecimiento de GSEA en las cuatro categorías de genes sinápticos (densidad postsináptica, densidad presináptica, zona activa y vesículas sinápticas), se puede comprobar que, a pesar de que, en general, los genes de la densidad postsináptica se hallan subexpresados en esquizofrenia, un grupo de ellos están altamente sobreexpresados, ocupando las posiciones más extremas en la figura, algo que no

ocurre de forma tan destacada con los otros tres grupos de genes de sinapsis (**figura 5.11**). De hecho, este grupo es el que se encontraba enriquecido el módulo C38.

Un total de 20 genes de los 628 del módulo C38 se encuentran en el set de densidad post sináptica de SynaptomeDB, y de ellos, 4 forman parte del 1% más sobreexpresado (223 genes que son el 1% de 22386, el total de genes expresados). Esto explica, claramente, que el pico de genes sobreexpresados en densidad postsináptica es explicable, al menos en parte, por el módulo C38 (valor P del test $\chi^2 - \text{Yates} = 1 \times 10^{-13}$).

Por otro lado, 7 genes del módulo sobreexpresado, y enriquecido en aquellos de astrocitos, M10 (*APOE*, *ATP1A2*, *GFAP*, *LRIG1*, *NTRK2*, *PMP2* y *SLC1A2*) albergan mutaciones en 15 casos y 3 genes (*GLUD1*, *ATP1A2* y *SLC1A2*) en 3 controles (**tabla 4.15**). Uno de estos genes, *GFAP*, cuya expresión se ha demostrado aumentada anteriormente en esquizofrenia (Catts y cols., 2014), se encuentra mutado en 3 casos (1 variante exónica y 2 mutaciones en la región 3'UTR) y ningún control. Una de las variantes UTR, rs567636153, se ubica en un lugar de unión de miRNA expresado en cerebro y descrito recientemente (Boudreau y col., 2014). El individuo con esta variante (A55), además, presenta el valor de expresión más alto de los 46 estudiados, casos y controles, en el análisis de variación genética (**figura 5.11**) lo que sugiere una posible contribución de la variante genética en la regulación positiva de este gen debida a la pérdida de eficiencia en la unión del miRNA al ARN transcrito y la consecuente reducción de la degradación de este ARN como mecanismo de regulación.

Figura 5. 11. Histograma de los valores de expresión del gen GFAP en las muestras analizadas en el estudio de variantes en transcriptoma de giro cingulado.



La variante rs57474185, por otra parte, provoca un cambio aminoacídico de prolina a leucina en la proteína, aunque ha sido descrita como "benigna" en *PolyPhen* y está presente a una frecuencia de 0,72% en la base de datos del ESP. Entre los otros genes mutados, otro caso es el de una variante exónica (rs769452) en *APOE*, clasificada como "probablemente dañina" en *PolyPhen*. Este gen ha sido implicado previamente en la actividad cerebral (Filippini y cols., 2009) y ampliamente estudiado en el ámbito de la esquizofrenia con resultados contradictorios (Martorell y cols., 2008; Tovilla-Zárate y cols., 2009). *PMP2* presentó una delección de 12 pb en la región UTR3' en un caso de esquizofrenia. La expresión alterada de *PMP2*, ya ha sido descrita previamente en corteza cingulada anterior (Dracheva y cols., 2006), aunque en el presente estudio no forma parte de los genes DE. El gen *NTRK2*, que codifica para un miembro de la familia de quinasas del receptor neurotrófico de tirosina, presentó una mutación en la región UTR3' en uno de los casos. *NTRK2* es el receptor de BDNF, y la interacción de ambos ha sido ampliamente estudiada en relación a esquizofrenia y otros trastornos psiquiátricos (Alonso y cols., 2008; Correia y cols., 2010; Juhasz y cols., 2011; Lin y cols., 2013b). En el estudio de las variantes en GWAS de esquizofrenia relativas a genes de astrocitos (Goudriaan y cols., 2013), el gen *NTRK2*, que es el único presente en los sets allí descritos por restringirse a aquellos genes únicamente expresados en un tipo celular, ya se encuentra asociado a nivel de gen (P corregido = 0,033). Este gen se encuentra en el set de genes de astrocitos enriquecidos en función de "transducción de señal". Curiosamente, el gen *NPAS3*, estudiado en la presente tesis, es otro de los que se asocian significativamente a esquizofrenia en el total de genes seleccionado en el estudio de Goudriaan y cols. (P corregido = 0,0032). *NTRK2* es uno de los genes diana de miRNAs que se han visto enriquecidos en CNV en esquizofrenia en un estudio reciente (Warnica y cols., 2015). Este gen también se ha visto relacionado, recientemente, en el estudio de asociación con intento de suicidio (Sokolowski y cols., 2015). La mayoría de las variantes presentes en genes de este módulo se encuentran en las regiones 3'UTR de los genes *SLC1A2* (5 en casos y una en controles) y *ATPIA2* (3 en casos y una en controles). El primero es un gen que codifica para un transportador de glutamato en células gliales, de importancia clave en procesos sinápticos. La variación genética en este gen ha sido recientemente implicada en esquizofrenia y otros trastornos neuropsiquiátricos, tanto en mutaciones raras de mayor efecto (Fiorentino y cols., 2015) como en variación común (Zhang y cols., 2015). Respecto a *ATPIA2*, miembro de la familia de ATPasas transportadoras de cationes, la variación genética en este gen se ha visto asociada a canalopatías relacionadas con migraña o epilepsia (Oh y cols., 2015).

Finalmente, mediante el análisis de varianza (ANOVA de tres vías) se demostró que las variantes descritas en los genes de astrocitos, que mostraban un enriquecimiento significativo en casos frente a controles (**tabla 4.15**), contribuían significativamente a las diferencias de expresión entre casos y controles (**tabla 4.18**). Estos resultados consolidan las hipótesis que

sugieren un mecanismo genético de susceptibilidad a esquizofrenia relacionado con astrocitos (Goudriaan y cols., 2013) y representan, en nuestro conocimiento, la primera evidencia que implica directamente la influencia en los estudios de expresión con las mismas muestras analizadas.

Mediante la comparación de los datos de expresión normalizando las diferencias intergénicas se ha podido determinar que la contribución significativa de las variantes en genes subexpresados de los módulos M1 y M2 (Hawrylycz y cols., 2012) era una asociación indirecta producto de un escenario en el que la mayor parte de las variantes se encontraban presentes en genes altamente expresados, como se había descrito anteriormente. No obstante, no fue el caso para las variantes de los módulos sobreexpresados M10 y C38, donde la contribución de la variable "presencia/ausencia de variante" era significativa incluso tras la supresión de las diferencias de expresión intergénicas ($P = 0,011$). Por tanto, el comentado caso de la variante rs567636153 en la muestra A55 no es una excepción, y con estos resultados se evidencia la implicación directa de la variación genética rara en genes de astrocitos como factores de riesgo a esquizofrenia mediante la alteración al alza de los niveles de expresión de los mismos.







6. CONCLUSIONES

1. El análisis de variantes en regiones reguladoras de la expresión de genes candidatos escogidos en base a patrones extremos de conservación (*TCF4*) o aceleración genética (*NPAS3*) ha demostrado no ser una buena estrategia en la búsqueda de variantes de susceptibilidad a esquizofrenia, sugiriendo la posibilidad de que la variación genética funcional en las mismas provoque fenotipos todavía más severos que el que caracteriza a esta enfermedad.
2. El análisis sistemático de regiones promotoras de varios genes asociados a esquizofrenia a raíz de estudios de GWAS o de CNV reveló una acumulación de variantes raras funcionales en casos esquizofrénicos respecto a controles, confirmando la validez de la hipótesis variante rara - enfermedad común en el contexto de las regiones reguladoras del genoma, y apoyando la necesidad de un mayor número de estudios de resecuenciación dirigida en estas regiones en detrimento de los de secuenciación de exomas.
3. El análisis de secuencias reguladoras de la expresión génica, mediante la utilización de anotación funcional específica de tejido disponible en bases de datos como Roadmap, Encode o BrainSpan, ha demostrado ser una buena aproximación para el estudio de variantes de susceptibilidad a esquizofrenia, sirviendo como precedente para la interpretación de futuros estudios de secuenciación de genoma completo en esquizofrenia.
4. El estudio de los estados de la cromatina descritos para tejido cerebral adulto y fetal a lo largo de las regiones analizadas ha servido para la identificación de regiones represoras Polycomb en los promotores de *LHX1* y *HNF1B*, cuya variación rara funcional confiere un riesgo a esquizofrenia similar o superior al de variantes muy penetrantes como CNV u otras codificantes disruptivas, respaldando la hipótesis inicial de este estudio que postulaba la existencia de variantes reguladoras de alto riesgo a esquizofrenia sometidas a selección purificadora.
5. Aunque estos resultados necesitan una confirmación en estudios independientes para su posterior validación, *a priori* constituyen una evidencia muy relevante de las alteraciones en los mecanismos epigenéticos de control de la expresión génica provocados por la presencia de variación genética reguladora en esquizofrenia, principalmente durante el neurodesarrollo temprano.
6. El análisis de expresión génica en muestras *post mortem* de giro cingulado en esquizofrenia revela una subexpresión y sobreexpresión de genes específicos de líneas celulares neuronales y de astrocitos, respectivamente, quedando demostrado, además, que estas alteraciones no son fruto de la medicación antipsicótica administrada a lo largo de la vida de los enfermos.

7. El análisis de variación genética rara en transcriptoma de giro cingulado demuestra un enriquecimiento significativo de las mismas en casos respecto a controles en aquellos genes pertenecientes a módulos sobreexpresados en esquizofrenia, relativos a astrocitos, y que explican al menos en parte, las diferencias de expresión observadas entre casos y controles en este grupo de genes. Este hallazgo establece la primera evidencia de la influencia directa de variantes genéticas de transcriptoma cerebral en las diferencias de expresión en genes de astrocitos, y constituye una demostración *in situ* de la importancia de la regulación de la expresión en el contexto de la esquizofrenia.
8. Tanto el análisis de variantes en ADN como en transcriptoma cerebral sugieren la importancia mayúscula de los procesos de regulación de la expresión génica en etapas tempranas del neurodesarrollo, en las que la variación genética rara ejerce una influencia muy relevante en el riesgo asociado a esquizofrenia.







7. BIBLIOGRAFÍA

1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), 56-65.

1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319), 1061-1073.

Abidin, S. Z., Tan, E. L., Chan, S. C., Jaafar, A., Lee, A. X., Hamid, M. H. N. A., ... & Lim, S. Y. (2015). DRD and GRIN2B polymorphisms and their association with the development of impulse control behaviour among Malaysian Parkinson's disease patients. *BMC neurology*, 15(1), 1.

Abramson, J. H. (2011). WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations*, 8(1), 1.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature methods*, 7(4), 248-249.

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L. A., & Rubin, E. M. (2007). Deletion of ultraconserved elements yields viable mice. *PLoS Biol*, 5(9), e234.

Akbadian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., Crawford, G. E., ... & Mill, J. (2015). The PsychENCODE project. *Nature neuroscience*, 18(12), 1707-1712.

Albanna, A., Choudhry, Z., Harvey, P. O., Fathalli, F., Cassidy, C., Sengupta, S. M., ... & Joobar, R. (2014). TCF4 gene polymorphism and cognitive performance in patients with first episode psychosis. *Schizophrenia research*, 152(1), 124-129.

Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*.

Allan, C. L., Cardno, A. G., Rijdsdijk, F. V., Kalidindi, S., Farmer, A., Murray, R. M., & McGuffin, P. (2009). Twin study of illness history variables in psychosis. *Schizophrenia research*, 115(2), 237-244.

Allardyce, J., Suppes, T., & Van Os, J. (2007). Dimensions and the psychosis phenotype. *International journal of methods in psychiatric research*, 16(S1), S34-S40.

Allen, N. C., Bagade, S., McQueen, M. B., Ioannidis, J. P., Kavvoura, F. K., Khoury, M. J., ... & Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nature genetics*, 40(7), 827-834.

- Alonso, P., Gratacòs, M., Menchón, J. M., Saiz-Ruiz, J., Segalàs, C., Baca-García, E., ... & Pérez, M. (2008). Extensive genotyping of the BDNF and NTRK2 genes define protective haplotypes against obsessive-compulsive disorder. *Biological psychiatry*, 63(6), 619-628.
- Altar, C. A., Jurata, L. W., Charles, V., Lemire, A., Liu, P., Bukhman, Y., ... & Knable, M. B. (2005). Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts. *Biological psychiatry*, 58(2), 85-96.
- Ament, S. A., Szelinger, S., Glusman, G., Ashworth, J., Hou, L., Akula, N., ... & Stittrich, A. B. (2015). Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proceedings of the National Academy of Sciences*, 112(11), 3576-3581.
- Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., ... & Colleaux, L. (2007). Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *The American Journal of Human Genetics*, 80(5), 988-993.
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology*, 11(10), R106.
- Anders, S., Pyl, P. T., & Huber, W. (2014). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, btu638.
- Andreasen, N. C. (2007). DSM and the death of phenomenology in America: an example of unintended consequences. *Schizophrenia bulletin*, 33(1), 108-112.
- Andreasen, N. C., & Pierson, R. (2008). The role of the cerebellum in schizophrenia. *Biological psychiatry*, 64(2), 81-88.
- Andreasen, N. C., Nopoulos, P., O'Leary, D. S., Miller, D. D., Wassink, T., & Flaum, M. (1999). Defining the phenotype of schizophrenia: cognitive dysmetria and its neural mechanisms. *Biological psychiatry*, 46(7), 908-920.
- Ardlie, K. G., Deluca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., ... & Ward, L. D. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235), 648-660.
- Ascano, M., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J. D., Corcoran, D. L., ... & Williams, Z. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature*, 492(7429), 382-386.

Bibliografía

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29.
- Ashe, P. C., Berry, M. D., & Boulton, A. A. (2001). Schizophrenia, a neurodegenerative disorder with neurodevelopmental antecedents. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 25(4), 691-707.
- Askland, K., Read, C., O'Connell, C., & Moore, J. H. (2012). Ion channels and schizophrenia: a gene set-based analytic approach to GWAS data for biological hypothesis testing. *Human genetics*, 131(3), 373-391.
- Asthana, S., Roytberg, M., Stamatoyannopoulos, J., & Sunyaev, S. (2007). Analysis of sequence conservation at nucleotide resolution. *PLoS Comput Biol*, 3(12), e254.
- Aston, C., Jiang, L., & Sokolov, B. P. (2004). Microarray analysis of postmortem temporal cortex from patients with schizophrenia. *Journal of neuroscience research*, 77(6), 858-866.
- Ayuso-Mateos, J. L., Gutierrez-Recacha, P., Haro, J. M., & Chisholm, D. (2006). Estimating the prevalence of schizophrenia in Spain using a disease model. *Schizophrenia research*, 86(1), 194-201.
- Bakewell, M. A., Shi, P., & Zhang, J. (2007). More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proceedings of the National Academy of Sciences*, 104(18), 7489-7494.
- Barratt, B. J., Payne, F., Rance, H. E., Nutland, S., Todd, J. A., & Clayton, D. G. (2002). Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Annals of human genetics*, 66(5-6), 393-405.
- Barrett, J. C., Fry, B., Maller, J. D. M. J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263-265.
- Barry, G., Briggs, J. A., Vanichkina, D. P., Poth, E. M., Beveridge, N. J., Ratnu, V. S., ... & Nakagawa, S. (2014). The long non-coding RNA GomaFu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Molecular psychiatry*, 19(4), 486-494.
- Basmanav, F. B., Forstner, A. J., Fier, H., Herms, S., Meier, S., Degenhardt, F., ... & Witt, S. H. (2015). Investigation of the role of TCF4 rare sequence variants in schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 168(5), 354-362.

- Bassett, A. S., Chow, E. W., & Weksberg, R. (2000). Chromosomal abnormalities and schizophrenia. *American journal of medical genetics*, 97(1), 45-51.
- Bayés, À., van de Lagemaat, L. N., Collins, M. O., Croning, M. D., Whittle, I. R., Choudhary, J. S., & Grant, S. G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nature neuroscience*, 14(1), 19-21.
- Bedont, J. L., LeGates, T. A., Slat, E. A., Byerly, M. S., Wang, H., Hu, J., ... & Hattar, S. (2014). Lhx1 controls terminal differentiation and circadian function of the suprachiasmatic nucleus. *Cell reports*, 7(3), 609-622.
- Begemann, M., Grube, S., Papiol, S., Malzahn, D., Krampe, H., Ribbe, K., ... & Hannke, K. (2010). Modification of cognitive performance in schizophrenia by complexin 2 gene polymorphisms. *Archives of general psychiatry*, 67(9), 879-888.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, 304(5675), 1321-1325.
- Bencherif, M., Stachowiak, M. K., Kucinski, A. J., & Lippiello, P. M. (2012). Alpha7 nicotinic cholinergic neuromodulation may reconcile multiple neurotransmitter hypotheses of schizophrenia. *Medical hypotheses*, 78(5), 594-600.
- Bender, H. U., Almashanu, S., Steel, G., Hu, C. A., Lin, W. W., Willis, A., ... & Valle, D. (2005). Functional consequences of PRODH missense mutations. *The American Journal of Human Genetics*, 76(3), 409-420.
- Benes, F. M., Lim, B., Matzilevich, D., Subburaju, S., & Walsh, J. P. (2008). Circuitry-based gene expression profiles in GABA cells of the trisynaptic pathway in schizophrenics versus bipolars. *Proceedings of the National Academy of Sciences*, 105(52), 20935-20940.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188.
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural brain research*, 125(1), 279-284.
- Benson, D. L., Huntsman, M. M., & Jones, E. G. (1994). Activity-dependent changes in GAD and preprotachykinin mRNAs in visual cortex of adult monkeys. *Cerebral Cortex*, 4(1), 40-51.

Bibliografía

Bernstein, H. G., Steiner, J., & Bogerts, B. (2009). Glial cells in schizophrenia: pathophysiological significance and possible consequences for therapy.

Bernstein, H. G., Steiner, J., Guest, P. C., Dobrowolny, H., & Bogerts, B. (2015). Glial cells as key players in schizophrenia pathology: recent insights and concepts of therapy. *Schizophrenia research*, 161(1), 4-18.

Bertolino, A., Caforio, G., Blasi, G., De Candia, M., Latorre, V., Petruzzella, V., ... & Mattay, V. S. (2014). Interaction of COMT Val108/158 Met genotype and olanzapine treatment on prefrontal cortical function in patients with schizophrenia. *American Journal of Psychiatry*.

Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A., & Cairns, M. J. (2010). Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Molecular psychiatry*, 15(12), 1176-1189.

Bird, C. P., Stranger, B. E., Liu, M., Thomas, D. J., Ingle, C. E., Beazley, C., ... & Dermitzakis, E. T. (2007). Fast-evolving noncoding sequences in the human genome. *Genome Biol*, 8(6), R118.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., ... & Kuehn, M. S. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), 799-816.

Bitanhirwe, B. K., & Woo, T. U. W. (2011). Oxidative stress in schizophrenia: an integrated approach. *Neuroscience & Biobehavioral Reviews*, 35(3), 878-893.

Blackledge, N. P., & Klose, R. (2011). CpG island chromatin: a platform for gene regulation. *Epigenetics*, 6(2), 147-152.

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J., & Klose, R. J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Molecular cell*, 38(2), 179-190.

Blackwood, D. H. R., Fordyce, A., Walker, M. T., Clair, D. S., Porteous, D. J., & Muir, W. J. (2001). Schizophrenia and affective disorders— cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *The American Journal of Human Genetics*, 69(2), 428-433.

Blangero, J. (2004). Localization and identification of human quantitative trait loci: king harvest has surely come. *Current opinion in genetics & development*, 14(3), 233-240.

- Bloomfield, P. S., Selvaraj, S., Veronese, M., Rizzo, G., Bertoldo, A., Owen, D. R., ... & McGuire, P. (2015). Microglial Activity in People at Ultra High Risk of Psychosis and in Schizophrenia: An [11C] PBR28 PET Brain Imaging Study. *American Journal of Psychiatry*.
- Blouin, J. L., Dombroski, B. A., Nath, S. K., Lasseter, V. K., Wolyniec, P. S., Nestadt, G., ... & Lamacz, M. (1998). Schizophrenia susceptibility *loci* on chromosomes 13q32 and 8p21. *Nature genetics*, 20(1), 70-73.
- Boksa, P. (2010). Effects of prenatal infection on brain development and behavior: a review of findings from animal models. *Brain, behavior, and immunity*, 24(6), 881-897.
- Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature genetics*, 33, 228-237.
- Boudreau, R. L., Jiang, P., Gilmore, B. L., Spengler, R. M., Tirabassi, R., Nelson, J. A., ... & Davidson, B. L. (2014). Transcriptome-wide discovery of microRNA binding sites in human brain. *Neuron*, 81(2), 294-305.
- Bowden, N. A., Scott, R. J., & Tooney, P. A. (2008). Altered gene expression in the superior temporal gyrus in schizophrenia. *BMC genomics*, 9(1), 199.
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., ... & Cherry, J. M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*, 22(9), 1790-1797.
- Braff, D. L., Ryan, J., Rissling, A. J., & Carpenter, W. T. (2013). Lack of use in the literature from the last 20 years supports dropping traditional schizophrenia subtypes from DSM-5 and ICD-11. *Schizophrenia bulletin*, sbt068.
- Brennand, K. J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., ... & McCarthy, S. (2011). Modelling schizophrenia using human induced pluripotent stem cells. *Nature*, 473(7346), 221-225.
- Brody, A. L., Mandelkern, M. A., Jarvik, M. E., Lee, G. S., Smith, E. C., Huang, J. C., ... & London, E. D. (2004). Differences between smokers and nonsmokers in regional gray matter volumes and densities. *Biological psychiatry*, 55(1), 77-84.
- Bromet, E. J., Kotov, R., Fochtmann, L. J., Carlson, G. A., Tanenberg-Karant, M., Ruggero, C., & Chang, S. W. (2011). Diagnostic shifts during the decade following first admission for psychosis. *American journal of psychiatry*, 168(11), 1186-1194.

Bibliografía

Brown, A. S., Begg, M. D., Gravenstein, S., Schaefer, C. A., Wyatt, R. J., Bresnahan, M., ...& Susser, E. S. (2004a). Serologic evidence of prenatal influenza in the etiology of schizophrenia. *Archives of general psychiatry*, 61(8), 774-780.

Brown, A. S., Bottiglieri, T., Schaefer, C. A., Quesenberry, C. P., Liu, L., Bresnahan, M., & Susser, E. S. (2007). Elevated prenatal homocysteine levels as a risk factor for schizophrenia. *Archives of General Psychiatry*, 64(1), 31-39.

Brown, A. S., Cohen, P., Greenwald, S., & Susser, E. (2014). Nonaffective psychosis after prenatal exposure to rubella. *American Journal of Psychiatry*.

Brown, A. S., Hooton, J., Schaefer, C. A., Zhang, H., Petkova, E., Babulas, V., ...& Susser, E. S. (2004b). Elevated maternal interleukin-8 levels and risk of schizophrenia in adult offspring. *American Journal of Psychiatry*.

Brown, A. S., Schaefer, C. A., Quesenberry Jr, C. P., Liu, L., Babulas, V. P., & Susser, E. S. (2005). Maternal exposure to toxoplasmosis and risk of schizophrenia in adult offspring. *American Journal of Psychiatry*, 162(4), 767-773.

Brown, S. (1997). Excess mortality of schizophrenia. A meta-analysis. *The British Journal of Psychiatry*, 171(6), 502-508.

Bruneau, E. G., McCullumsmith, R. E., Haroutunian, V., Davis, K. L., & Meador-Woodruff, J. H. (2005). Increased expression of glutaminase and glutamine synthetase mRNA in the thalamus in schizophrenia. *Schizophrenia research*, 75(1), 27-34.

Brunskill, E. W., Ehrman, L. A., Williams, M. T., Klanke, J., Hammer, D., Schaefer, T. L., ... & Vorhees, C. V. (2005). Abnormal neurodevelopment, neurosignaling and behaviour in Npas3-deficient mice. *European Journal of Neuroscience*, 22(6), 1265-1276.

Brunskill, E. W., Witte, D. P., Shreiner, A. B., & Potter, S. S. (1999). Characterization of npas3, a novel basic helix-loop-helix PAS gene expressed in the developing mouse nervous system. *Mechanisms of development*, 88(2), 237-241.

Bruton, C. J., Crow, T. J., Frith, C. D., Johnstone, E. C., Owens, D. G. C., & Roberts, G. W. (1990). Schizophrenia and the brain: a prospective clinico-neuropathological study. *Psychological medicine*, 20(02), 285-304.

Brzózka, M. M., Radyushkin, K., Wichert, S. P., Ehrenreich, H., & Rossner, M. J. (2010). Cognitive and sensorimotor gating impairments in transgenic mice overexpressing the schizophrenia susceptibility gene Tcf4 in the brain. *Biological psychiatry*, 68(1), 33-40.

- Buka, S. L., Tsuang, M. T., Torrey, E. F., Klebanoff, M. A., Bernstein, D., & Yolken, R. H. (2001). Maternal infections and subsequent psychosis among offspring. *Archives of general psychiatry*, 58(11), 1032-1037.
- Bundy, H., Stahl, D., & MacCabe, J. H. (2011). A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. *Acta Psychiatrica Scandinavica*, 123(2), 98-106.
- Bush, E. C., & Lahn, B. T. (2008). A genome-wide screen for noncoding elements important in primate evolution. *BMC evolutionary biology*, 8(1), 17.
- Bushe, C., & Holt, R. (2004). Prevalence of diabetes and impaired glucose tolerance in patients with schizophrenia. *The British Journal of Psychiatry*, 184(47), s67-s71.
- Busse, S., Busse, M., Schiltz, K., Biela, H., Gos, T., Brisch, R., ... & Bernstein, H. G. (2012). Different distribution patterns of lymphocytes and microglia in the hippocampus of patients with residual versus paranoid schizophrenia: further evidence for disease course-related immune alterations? *Brain, behavior, and immunity*, 26(8), 1273-1279.
- Butler, A. A., Webb, W. M., & Lubin, F. D. (2016). Regulatory RNAs and control of epigenetic mechanisms: expectations for cognition and cognitive dysfunction. *Epigenomics*, 8(1), 135-151.
- Byers, D. (2008). Components of phenotypic variance. *Nature Education*, 1(1), 161.
- Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., ... & Barres, B. A. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *The Journal of Neuroscience*, 28(1), 264-278.
- Cannon, M., Jones, P. B., & Murray, R. M. (2002). Obstetric complications and schizophrenia: historical and meta-analytic review. *American Journal of Psychiatry*.
- Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L., & Pollard, K. S. (2013). Many human accelerated regions are developmental enhancers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1632), 20130025.
- Cardno, A. G., & Gottesman, I. I. (2000). Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *American journal of medical genetics*, 97(1), 12-17.
- Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics*, 2(2), 91-99.

Bibliografía

- Carlson, C. S., Eberle, M. A., Rieder, M. J., Yi, Q., Kruglyak, L., & Nickerson, D. A. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *The American Journal of Human Genetics*, 74(1), 106-120.
- Carlsson, A. (1977). Does dopamine play a role in schizophrenia?. *Psychological medicine*, 7(04), 583-597.
- Carlsson, A., & Lindqvist, M. (1963). Effect of chlorpromazine or haloperidol on formation of 3-methoxytyramine and normetanephrine in mouse brain. *Acta pharmacologica et toxicologica*, 20(2), 140-144.
- Carpenter, W. T., & Tandon, R. (2013). Psychotic disorders in DSM-5: summary of changes. *Asian journal of psychiatry*, 6(3), 266-268.
- Carrera, N., Arrojo, M., Sanjuán, J., Ramos-Ríos, R., Paz, E., Suárez-Rama, J. J., ... & Rivero, O. (2012). Association study of nonsynonymous single nucleotide polymorphisms in schizophrenia. *Biological psychiatry*, 71(2), 169-177.
- Carroll, L. S., & Owen, M. J. (2009). Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome medicine*, 1(10), 1-7.
- Carter, C. J. (2006). Schizophrenia susceptibility genes converge on interlinked pathways related to glutamatergic transmission and long-term potentiation, oxidative stress and oligodendrocyte viability. *Schizophrenia research*, 86(1), 1-14.
- Cassidy, C. M., Lepage, M., & Malla, A. (2014). Do motivation deficits in schizophrenia-spectrum disorders promote cannabis use? An investigation of behavioural response to natural rewards and drug cues. *Psychiatry research*, 215(3), 522-527.
- Catts, V. S., Wong, J., Fillman, S. G., Fung, S. J., & Weickert, C. S. (2014). Increased expression of astrocyte markers in schizophrenia: association with neuroinflammation. *Australian and New Zealand Journal of Psychiatry*, 48(8), 722-734.
- Cermolacce, M., Sass, L., & Parnas, J. (2010). What is bizarre in bizarre delusions? A critical review. *Schizophrenia bulletin*, sbq001.
- Chan, M. K., Tsang, T. M., Harris, L. W., Guest, P. C., Holmes, E., & Bahn, S. (2011). Evidence for disease and antipsychotic medication effects in post-mortem brain from schizophrenia patients. *Molecular psychiatry*, 16(12), 1189-1202.

- Chandley, M. J., Szebeni, K., Szebeni, A., Crawford, J., Stockmeier, C. A., Turecki, G., ...& Ordway, G. A. (2013). Gene expression deficits in pontine locus coeruleus astrocytes in men with major depressive disorder. *J Psychiatry Neurosci*, 38(4), 276-84.
- Chen, X., & Sullivan, P. F. (2003). Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *The pharmacogenomics journal*, 3(2), 77-96.
- Chen, X., Long, F., Cai, B., Chen, X., & Chen, G. (2015). A novel relationship for schizophrenia, bipolar and major depressive disorder Part 7: A hint from chromosome 7 high density association screen. *Behavioural brain research*, 293, 241-251.
- Cheung, I., Shulha, H. P., Jiang, Y., Matevossian, A., Wang, J., Weng, Z., & Akbarian, S. (2010). Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. *Proceedings of the National Academy of Sciences*, 107(19), 8824-8829.
- Chiaromonte, F., Weber, R. J., Roskin, K. M., Diekhans, M., Kent, W. J., & Haussler, D. (2003, January). The share of human genomic DNA under selection estimated from human–mouse genomic alignments. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 68, pp. 245-254). Cold Spring Harbor Laboratory Press.
- Chinwalla, A. T., Cook, L. L., Delehaunty, K. D., Fewell, G. A., Fulton, L. A., Fulton, R. S., ... & Miner, T. L. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), 520-562.
- Chiocchetti, A. G., Kopp, M., Waltes, R., Haslinger, D., Duketis, E., Jarczok, T. A., ... & Klauck, S. M. (2014). Variants of the CNTNAP2 5' promoter as risk factors for autism spectrum disorders: a genetic and functional approach. *Molecular psychiatry*.
- Choi, K. H., Elashoff, M., Higgs, B. W., Song, J., Kim, S., Sabunciyan, S., ... & Webster, M. J. (2008). Putative psychosis genes in the prefrontal cortex: combined analysis of gene expression microarrays. *BMC psychiatry*, 8(1), 87.
- Choi, K. H., Zepp, M. E., Higgs, B. W., Weickert, C. S., & Webster, M. J. (2009a). Expression profiles of schizophrenia susceptibility genes during human prefrontal cortical development. *Journal of psychiatry & neuroscience: JPN*, 34(6), 450.
- Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., ...& Nelson-Williams, C. (2009b). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences*, 106(45), 19096-19101.

- Chomette, D., Frain, M., Cereghini, S., Charnay, P., & Ghislain, J. (2006). Krox20 hindbrain cis-regulatory landscape: interplay between multiple long-range initiation and autoregulatory elements. *Development*, 133(7), 1253-1262.
- Chow, M. L., Pramparo, T., Winn, M. E., Barnes, C. C., Li, H. R., Weiss, L., ...& Courchesne, E. (2012). Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet*, 8(3), e1002592.
- Chun, S., Westmoreland, J. J., Bayazitov, I. T., Eddins, D., Pani, A. K., Smeyne, R. J., ...& Zakharenko, S. S. (2014). Specific disruption of thalamic inputs to the auditory cortex in schizophrenia models. *Science*, 344(6188), 1178-1182.
- Ciampi, L. (1980) Catamnestic long-term study on the course of life and aging of schizophrenics. *Schizophr Bull*, 6(4):606-618.
- Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, 11(6), 415-425.
- Clark, T. A., Sugnet, C. W., & Ares, M. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science*, 296(5569), 907-910.
- Clarke, M. C., Tanskanen, A., Huttunen, M., Whittaker, J. C., & Cannon, M. (2009). Evidence for an interaction between familial liability and prenatal exposure to infection in the causation of schizophrenia. *American Journal of Psychiatry*, 166(9), 1025-1030.
- Clarke, M., O'Callaghan, E. (2003). First-episode studies in schizophrenia. In *The Epidemiology of Schizophrenia* (ed. R. M. Murray, P. B. Jones, E. Susser, J. van Os and M. E. Cannon), pp. 148–166. Cambridge University Press :Cambridge.
- Cline, J., Braman, J. C., & Hogrefe, H. H. (1996). PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic acids research*, 24(18), 3546-3551.
- Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., ... & Robertson, A. J. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature methods*, 5(7), 613-619.
- Collins, A. L., Kim, Y., Bloom, R. J., Kelada, S. N., Sethupathy, P., & Sullivan, P. F. (2014). Transcriptional targets of the schizophrenia risk gene MIR137. *Translational psychiatry*, 4(7), e404.

- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., ... & Fitzgerald, T. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), 704-712.
- Cooper, G. M., Stone, E. A., Asimenos, G., Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome research*, 15(7), 901-913.
- Cooper, J. E., Kendell, R. E., Gurland, B. J., Sharpe, L., Copeland, J. R. M., & Simon, R. (1972). Psychiatric diagnosis in New York and London: A comparative study of mental hospital admissions. Institute of Psychiatry, Maudsley Monographs, No 20.
- Correia, C. T., Coutinho, A. M., Sequeira, A. F., Sousa, I. G., Lourenco Venda, L., Almeida, J. P., ... & Cochrane, L. (2010). Increased BDNF levels and NTRK2 gene association suggest a disruption of BDNF/TrkB signaling in autism. *Genes, Brain and behavior*, 9(7), 841-848.
- Costain, G., Ho, A., Crawley, A. P., Mikulis, D. J., Brzustowicz, L. M., Chow, E. W., & Bassett, A. S. (2010). Reduced gray matter in the anterior cingulate gyrus in familial schizophrenia: a preliminary report. *Schizophrenia research*, 122(1), 81-84.
- Costain, G., Lionel, A. C., Merico, D., Forsythe, P., Russell, K., Lowther, C., ... & Chow, E. W. (2013). Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *Human molecular genetics*, 22(22), 4485-4501.
- Costas, J. Estudios genéticos de asociación en enfermedades multifactoriales complejas. *Neurol Supl* 2005;1(3):44-50.
- Costas, J., Carrera, N., Domínguez, E., Vilella, E., Martorell, L., Valero, J., ... & Carracedo, Á. (2009). A common haplotype of DRD3 affected by recent positive selection is associated with protection from schizophrenia. *Human genetics*, 124(6), 607-613.
- Costas, J., Sanjuán, J., Ramos-Ríos, R., Paz, E., Agra, S., Ivorra, J. L., ... & Arrojo, M. (2011). Heterozygosity at catechol-O-methyltransferase Val158Met and schizophrenia: new data and meta-analysis. *Journal of psychiatric research*, 45(1), 7-14.
- Costas, J., Suárez-Rama, J. J., Carrera, N., Paz, E., Páramo, M., Agra, S., ... & Arrojo, M. (2013). Role of DISC1 Interacting Proteins in Schizophrenia Risk from Genome-Wide Analysis of Missense SNP. *Annals of human genetics*, 77(6), 504-512.
- Coyle, J. T. (2006). Glutamate and schizophrenia: beyond the dopamine hypothesis. *Cellular and molecular neurobiology*, 26(4-6), 363-382.

Bibliografía

Craddock, N., & Owen, M. J. (2010). The Kraepelinian dichotomy—going, going... but still not gone. *The British Journal of Psychiatry*, 196(2), 92-95.

Crespi, B., Summers, K., & Dorus, S. (2007). Adaptive evolution of genes underlying schizophrenia. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1627), 2801-2810.

Crespo-Facorro, B., Prieto, C., & Sainz, J. (2015). Schizophrenia gene expression profile reverted to normal levels by antipsychotics. *International Journal of Neuropsychopharmacology*, 18(4), pyu066.

Cropley, V. L., Fujita, M., Innis, R. B., & Nathan, P. J. (2006). Molecular imaging of the dopaminergic system and its association with human cognitive function. *Biological psychiatry*, 59(10), 898-907.

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013a). Genetic relationship between five psychiatric disorders estimated from genome-wide SNP. *Nature genetics*, 45(9), 984-994.

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013b). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, 381(9875), 1371-1379.

Crow, J. F. (2003). There's something curious about paternal-age effects. *Science*, 301(5633), 606-607.

Crow, T. J. (1995). Constraints on concepts of pathogenesis: Language and the speciation process as the key to the etiology of schizophrenia. *Archives of general psychiatry*, 52(12), 1011-1014.

Crow, T. J. (1997). Is schizophrenia the price that Homo sapiens pays for language?. *Schizophrenia research*, 28(2), 127-141.

Crow, T. J. (2008). The 'big bang' theory of the origin of psychosis and the faculty of language. *Schizophrenia research*, 102(1), 31-52.

Crowley, J. J., Hilliard, C. E., Kim, Y., Morgan, M. B., Lewis, L. R., Muzny, D. M., ... & Sullivan, P. F. (2013). Deep resequencing and association analysis of schizophrenia candidate genes. *Molecular psychiatry*, 18(2), 138.

- Cuevas-Esteban, J., Campayo, A., Gutiérrez-Galve, L., Gracia-García, P., & López-Antón, R. (2011). Fundamentos y hallazgos de la neuroimagen en la esquizofrenia: una actualización. *Rev Neurol*, 52, 27-36.
- Damadzić, R., Bigelow, L. B., Krimer, L. S., Goldenson, D. A., Saunders, R. C., Kleinman, J. E., & Herman, M. M. (2001). A quantitative immunohistochemical study of astrocytes in the entorhinal cortex in schizophrenia, bipolar disorder and major depression: absence of significant astrogliosis. *Brain research bulletin*, 55(5), 611-618.
- Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., ... & Licatalosi, D. D. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*, 146(2), 247-261.
- Davis, K. L., & Haroutunian, V. (2003). Global expression-profiling studies and oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *The Lancet*, 362(9386), 758.
- Davis, K. N., Tao, R., Li, C., Gao, Y., Gondré-Lewis, M. C., Lipska, B. K., ... & Kleinman, J. E. (2016). GAD2 Alternative Transcripts in the Human Prefrontal Cortex, and in Schizophrenia and Affective Disorders. *PloS one*, 11(2), e0148558.
- Davis, L. K., Maltman, N., Mosconi, M. W., Macmillan, C., Schmitt, L., Moore, K., ... & Cook, E. H. (2012). Rare inherited A2BP1 deletion in a proband with autism and developmental hemiparesis. *American Journal of Medical Genetics Part A*, 158(7), 1654-1661.
- de Jong, S., Boks, M. P., Fuller, T. F., Strengman, E., Janson, E., de Kovel, C. G., ... & Glenthøj, B. (2012). A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. *PloS one*, 7(6), e39498.
- de Pontual, L., Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., ... & Heron, D. (2009). Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. *Human mutation*, 30(4), 669-676.
- De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., ... & Singh, T. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526), 209-215.
- Degenhardt, F., Priebe, L., Meier, S., Lennertz, L., Streit, F., Witt, S. H., ... & Nenadic, I. (2013). Duplications in RB1CC1 are associated with schizophrenia; identification in large European sample sets. *Translational psychiatry*, 3(11), e326.

- Delalle, I., Kao, P., & Choi, J. (2014). Deregulated microRNA expression in biospecimens from patients diagnosed with schizophrenia and bipolar disorder as a disease biomarker. *Translational Neuroscience*, 5(3), 173-178.
- Dempster, E. L., Pidsley, R., Schalkwyk, L. C., Owens, S., Georgiades, A., Kane, F., ... & Murray, R. M. (2011). Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Human molecular genetics*, ddr416.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... & McKenna, A. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43(5), 491-498.
- Dermitzakis, E. T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, C., Deutsch, S., ... & Antonarakis, S. E. (2002). Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature*, 420(6915), 578-582.
- Dermitzakis, E. T., Reymond, A., Scamuffa, N., Ucla, C., Kirkness, E., Rossier, C., & Antonarakis, S. E. (2003). Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science*, 302(5647), 1033-1035.
- Devlin, B., & Risch, N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, 29(2), 311-322.
- Dioum, E. M., Rutter, J., Tuckerman, J. R., Gonzalez, G., Gilles-Gonzalez, M. A., & McKnight, S. L. (2002). NPAS2: a gas-responsive transcription factor. *Science*, 298(5602), 2385-2387.
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., ... & Xue, C. (2012). Landscape of transcription in human cells. *Nature*, 489(7414), 101-108.
- Do, K. Q., Trabesinger, A. H., Kirsten-Krüger, M., Lauer, C. J., Dydak, U., Hell, D., ... & Cuenod, M. (2000). Schizophrenia: glutathione deficit in cerebrospinal fluid and prefrontal cortex in vivo. *European Journal of Neuroscience*, 12(10), 3721-3728.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21.
- Doherty, J. L., & Owen, M. J. (2014). Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med*, 6(4), 29.
- Dong, X., Greven, M. C., Kundaje, A., Djebali, S., Brown, J. B., Cheng, C., ... & Weng, Z. (2012). Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol*, 13(9), R53.

- Dougherty, J. D., Schmidt, E. F., Nakajima, M., & Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic acids research*, 38(13), 4218-4230.
- Dowlati, Y., Herrmann, N., Swardfager, W., Liu, H., Sham, L., Reim, E. K., & Lanctôt, K. L. (2010). A meta-analysis of cytokines in major depression. *Biological psychiatry*, 67(5), 446-457.
- Doyle, J. P., Dougherty, J. D., Heiman, M., Schmidt, E. F., Stevens, T. R., Ma, G., ... & Gong, S. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell*, 135(4), 749-762.
- Dracheva, S., Davis, K. L., Chin, B., Woo, D. A., Schmeidler, J., & Haroutunian, V. (2006). Myelin-associated mRNA and protein expression deficits in the anterior cingulate cortex and hippocampus in elderly schizophrenia patients. *Neurobiology of disease*, 21(3), 531-540.
- Drevets, W. C., Savitz, J., & Trimble, M. (2008). The subgenual anterior cingulate cortex in mood disorders. *CNS spectrums*, 13(8), 663.
- Drews, C., Senkel, S., & Ryffel, G. U. (2011). The nephrogenic potential of the transcription factors *osr1*, *osr2*, *hnf1b*, *lhx1* and *pax8* assessed in *Xenopus* animal caps. *BMC developmental biology*, 11(1), 5.
- Duan, J., Shi, J., Fiorentino, A., Leites, C., Chen, X., Moy, W., ... & Freda, J. (2014). A rare functional noncoding variant at the GWAS-Implicated MIR137/MIR2682 locus might confer risk to schizophrenia and bipolar disorder. *The American Journal of Human Genetics*, 95(6), 744-753.
- Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Human heredity*, 66(2), 87-98.
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293), 52-64.
- Duret, L., & Galtier, N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual review of genomics and human genetics*, 10, 285-311.
- Dwight, Z., Palais, R., & Wittwer, C. T. (2011). uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics*, 27(7), 1019-1020.

Dwyer, S., Williams, H., Jones, I., Jones, L., Walters, J., Craddock, N., Owen, M.J., O'Donovan, M.C., 2011. Investigation of rare non-synonymous variants at ABCA13 in schizophrenia and bipolar disorder. *Mol Psychiatry* 16, 790-791.

Ecker, J. R., Bickmore, W. A., Barroso, I., Pritchard, J. K., Gilad, Y., & Segal, E. (2012). Genomics: ENCODE explained. *Nature*, 489(7414), 52-55.

Edwards, T. L., Wang, X., Chen, Q., Wormly, B., Riley, B., O'Neill, F. A., ... & Chen, X. (2008). Interaction between interleukin 3 and dystrobrevin-binding protein 1 in schizophrenia. *Schizophrenia research*, 106(2), 208-217.

Efron, B. and Tibshirani, J. (1993). An introduction to the bootstrap. Chapman & Hall

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., ... & Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences*, 98(12), 6917-6922.

Ek, M., Wicks, S., Svensson, A. C., Idring, S., & Dalman, C. (2014). Advancing paternal age and schizophrenia: the impact of delayed fatherhood. *Schizophrenia bulletin*, sbu154.

Elia, J., Glessner, J. T., Wang, K., Takahashi, N., Shtir, C. J., Hadley, D., ... & Lyon, G. J. (2012). Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nature genetics*, 44(1), 78-84.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., ... & Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418(6900), 869-872.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57-74.

Engstrom EJ. Clinical Psychiatry in Imperial Germany: A History of Psychiatric Practice. Ithaka, NY: Cornell University Press; 2004.

Erali, M., & Wittwer, C. T. (2010). High resolution melting analysis for gene scanning. *Methods*, 50(4), 250-261.

Erbel-Sieler, C., Dudley, C., Zhou, Y., Wu, X., Estill, S. J., Han, T., ... & McKnight, S. L. (2004). Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 101(37), 13648-13653.

- Eroglu, C., & Barres, B. A. (2010). Regulation of synaptic connectivity by glia. *Nature*, 468(7321), 223-231.
- Eskandari, M. R., Abdolmaleky, H. M., Zhou, J. R., & Thiagalingam, S. (2015). Reduced Risk of Cancer in Schizophrenia, a Bridge Toward Etio-Pathology and Therapy of Both Diseases. In *Epigenetics Territory and Cancer* (pp. 137-166). Springer Netherlands.
- Etherton, M. R., Blaiss, C. A., Powell, C. M., & Südhof, T. C. (2009). Mouse neurexin-1 α deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. *Proceedings of the National Academy of Sciences*, 106(42), 17998-18003.
- Evensen, S., Wisløff, T., Lystad, J. U., Bull, H., Ueland, T., & Falkum, E. (2015). Prevalence, Employment Rate, and Cost of Schizophrenia in a High-Income Welfare Society: A Population-Based Study Using Comprehensive Health and Welfare Registers. *Schizophreniabulletin*, sbv141.
- Falkai, P., Honer, W. G., David, S., Bogerts, B., Majtenyi, C., & Bayer, T. A. (1999). No evidence for astrogliosis in brains of schizophrenic patients. A post-mortem study. *Neuropathology and applied neurobiology*, 25(1), 48-53.
- Fang, F., Hodges, E., Molaro, A., Dean, M., Hannon, G. J., & Smith, A. D. (2012). Genomic landscape of human allele-specific DNA methylation. *Proceedings of the National Academy of Sciences*, 109(19), 7332-7337.
- Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., ... & Koseki, H. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *elife*, 1, e00205.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of general psychiatry*, 26(1), 57-63.
- Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature*, 447(7143), 433-440.

- Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., & Gloor, G. B. (2013). ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One*, 8(7), e67019.
- Ferreira, M. A., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., ... & Smoller, J. W. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature genetics*, 40(9), 1056-1058.
- Fields, R. D. (2005). Myelination: an overlooked mechanism of synaptic plasticity?. *The Neuroscientist*, 11(6), 528-531.
- Fields, R. D. (2008). White matter in learning, cognition and psychiatric disorders. *Trends in neurosciences*, 31(7), 361-370.
- Filippini, N., MacIntosh, B. J., Hough, M. G., Goodwin, G. M., Frisoni, G. B., Smith, S. M., ... & Mackay, C. E. (2009). Distinct patterns of brain activity in young carriers of the APOE-ε4 allele. *Proceedings of the National Academy of Sciences*, 106(17), 7209-7214.
- Fillman, S. G., Cloonan, N., Catts, V. S., Miller, L. C., Wong, J., McCrossin, T., ... & Weickert, C. S. (2013). Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Molecular psychiatry*, 18(2), 206-214.
- Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., ... & Ripke, S. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*, 47(11), 1228-1235.
- Fiorentino, A., Sharp, S. I., & McQuillin, A. (2015). Association of rare variation in the glutamate receptor gene SLC1A2 with susceptibility to bipolar disorder and schizophrenia. *European Journal of Human Genetics*, 23(9), 1200-1206.
- Folstein, S. E., & Rosen-Sheidley, B. (2001). Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Reviews Genetics*, 2(12), 943-955.
- Forrest, M., Chapman, R. M., Doyle, A. M., Tinsley, C. L., Waite, A., & Blake, D. J. (2012). Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome. *Human mutation*, 33(12), 1676-1686.
- Franco, S. J., Martinez-Garay, I., Gil-Sanz, C., Harkins-Perry, S. R., & Müller, U. (2011). Reelin regulates cadherin function via Dab1/Rap1 to control neuronal migration and lamination in the neocortex. *Neuron*, 69(3), 482-497.

- Frankish, A., Uszczynska, B., Ritchie, G. R., Gonzalez, J. M., Pervouchine, D., Petryszak, R., ...&Harrow, J. (2015). Comparison of GENCODE and RefSeq gene annotation and the impact of reference gene sets on variant effect prediction. *BMC genomics*, 16(Suppl 8), S2.
- Friedman, J. I., Vrijenhoek, T., Markx, S., Janssen, I. M., Van Der Vliet, W. A., Faas, B. H. W., ... & Davis, K. L. (2008). CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Molecular psychiatry*, 13(3), 261-266.
- Friedman, J. I., Wallenstein, S., Moshier, E., Parrella, M., White, L., Bowler, S., ...& Davis, K. L. (2010). The effects of hypertension and body mass index on cognition in schizophrenia. *American Journal of Psychiatry*.
- Friis, S., Melle, I., Johannessen, J. O., Røssberg, J. I., Barder, H. E., Evensen, J. H., ... & Larsen, T. K. (2015). Early Predictors of Ten-Year Course in First-Episode Psychosis. *Psychiatric Services*.
- Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., ...& Carrera, N. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature*, 506(7487), 179-184.
- Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., Matthews, R. G., ... & Rozen, R. (1995). A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase.
- Fu, W., O'Connor, T. D., Jun, G., Kang, H. M., Abecasis, G., Leal, S. M., ... & Nickerson, D. A. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*, 493(7431), 216-220.
- Fung, S. J., Webster, M. J., Sivagnanasundaram, S., Duncan, C., Elashoff, M., & Weickert, C. S. (2010). Expression of interneuron markers in the dorsolateral prefrontal cortex of the developing human and in schizophrenia. *American Journal of Psychiatry*.
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., ...&Liu-Cordero, S. N. (2002). The structure of haplotype blocks in the human genome. *Science*, 296(5576), 2225-2229.
- Ganguli, R., &Gubbi, A. (1997). Clinical and immunological characteristics of a subgroup of patients suffering from schizophrenia.

- Garbett, K., Ebert, P. J., Mitchell, A., Lintas, C., Manzi, B., Mirnics, K., & Persico, A. M. (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiology of disease*, 30(3), 303-311.
- Garey, L. (2010). When cortical development goes wrong: schizophrenia as a neurodevelopmental disease of microcircuits. *Journal of anatomy*, 217(4), 324-333.
- Gavin, D. P., & Sharma, R. P. (2010). Histone modifications, DNA methylation, and schizophrenia. *Neuroscience & Biobehavioral Reviews*, 34(6), 882-888.
- Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., ... & Min, R. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489(7414), 91-100.
- Geyer, M. A., & Vollenweider, F. X. (2008). Serotonin research: contributions to understanding psychoses. *Trends in pharmacological sciences*, 29(9), 445-453.
- Gillberg, C. (1998). Chromosomal disorders and autism. *Journal of autism and developmental disorders*, 28(5), 415-425.
- Gilles-Gonzalez, M. A., & Gonzalez, G. (2005). Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. *Journal of inorganic biochemistry*, 99(1), 1-22.
- Gilman, S. R., Chang, J., Xu, B., Bawa, T. S., Gogos, J. A., Karayiorgou, M., & Vitkup, D. (2012). Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nature neuroscience*, 15(12), 1723-1728.
- Girard, S. L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., ... & Thibodeau, P. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics*, 43(9), 860-863.
- Girirajan, S., Campbell, C. D., & Eichler, E. E. (2011). Human copy number variation and complex genetic disease. *Annual review of genetics*, 45, 203-226.
- Girirajan, S., Rosenfeld, J. A., Cooper, G. M., Antonacci, F., Siswara, P., Itsara, A., ... & Mefford, H. C. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nature genetics*, 42(3), 203-209.
- Gkountela, S., Zhang, K. X., Shafiq, T. A., Liao, W. W., Hargan-Calvopiña, J., Chen, P. Y., & Clark, A. T. (2015). DNA demethylation dynamics in the human prenatal germline. *Cell*, 161(6), 1425-1436.

- Glatt, S. J., Faraone, S. V., & Tsuang, M. T. (2003). Association between a functional catechol O-methyltransferase gene polymorphism and schizophrenia: meta-analysis of case-control and family-based studies. *American Journal of Psychiatry*.
- Goldar, J. C., Starkstein, S. E., & Hodgkiss, A. (1995). Karl Ludwig Kahlbaum's concept of catatonia. *History of psychiatry*, 6(22), 201-207.
- Goldman, A. L., Pezawas, L., Mattay, V. S., Fischl, B., Verchinski, B. A., Zolnick, B., ... & Meyer-Lindenberg, A. (2008). Heritability of brain morphology related to schizophrenia: a large-scale automated magnetic resonance imaging segmentation study. *Biological psychiatry*, 63(5), 475-483.
- Goldstein, D. B., & Chikhi, L. (2002). Human migrations and population structure: what we know and why it matters. *Annual review of genomics and human genetics*, 3(1), 129-152.
- Goldstein, D. B., Ahmadi, K. R., Weale, M. E., & Wood, N. W. (2003). Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *TRENDS in Genetics*, 19(11), 615-622.
- Golzio, C., Willer, J., Talkowski, M. E., Oh, E. C., Taniguchi, Y., Jacquemont, S., ... & Kamiya, A. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature*, 485(7398), 363-367.
- González-Penas, J., Arrojo, M., Páramo, M., Paz, E., Agra, S., Ramos-Ríos, R., ... & Costas, J. (2014). Absence of low frequency variants associated with schizophrenia at the ultraconserved non-coding region of TCF4. *Psychiatry research*, 215(1), 255.
- González-Peñas, J., Arrojo, M., Paz, E., Brenlla, J., Páramo, M., & Costas, J. (2015). Cumulative role of rare and common putative functional genetic variants at NPAS3 in schizophrenia susceptibility. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 168(7), 528-535.
- Gorlov, I. P., Gorlova, O. Y., Sunyaev, S. R., Spitz, M. R., & Amos, C. I. (2008). Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *The American Journal of Human Genetics*, 82(1), 100-112.
- Gottesman, I. I. & Shields, J. (1982) Schizophrenia: The epigenetic puzzle.
- Gottesman, I. I. (1991). *Schizophrenia genesis: The origins of madness*. WH Freeman/Times Books/Henry Holt & Co.

Bibliografía

- Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of Psychiatry*, 160(4), 636-645.
- Gottesman, I. I., & Shields, J. (1967). A polygenic theory of schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 58(1), 199.
- Goudriaan, A., De Leeuw, C., Ripke, S., Hultman, C. M., Sklar, P., Sullivan, P. F., ... & Verheijen, M. H. (2013). Specific glial functions contribute to schizophrenia susceptibility. *Schizophrenia bulletin*, sbt109.
- Graeber, M. B. (2010). Changing face of microglia. *Science*, 330(6005), 783-788.
- Graham, R., Liew, M., Meadows, C., Lyon, E., & Wittwer, C. T. (2005). Distinguishing different DNA heterozygotes by high-resolution melting. *Clinical Chemistry*, 51(7), 1295-1298.
- Grau, D. J., Chapman, B. A., Garlick, J. D., Borowsky, M., Francis, N. J., & Kingston, R. E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes & development*, 25(20), 2210-2221.
- Green, E. K., Grozeva, D., Jones, I., Jones, L., Kirov, G., Caesar, S., ... & Hamshere, M. L. (2010). The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Molecular psychiatry*, 15(10), 1016-1022.
- Greenwood, T. A., Braff, D. L., Light, G. A., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., ... & Mintz, J. (2007). Initial heritability analyses of endophenotypic measures for schizophrenia: the consortium on the genetics of schizophrenia. *Archives of general psychiatry*, 64(11), 1242-1250.
- Grinshpoon, A., Barchana, M., Ponizovsky, A., Lipshitz, I., Nahon, D., Tal, O., ... & Levay, I. (2005). Cancer in schizophrenia: is the risk higher or lower?. *Schizophrenia research*, 73(2), 333-341.
- Grossman, A. R., Bhaya, D., & He, Q. (2001). Tracking the light environment by cyanobacteria and the dynamic nature of light harvesting. *Journal of Biological Chemistry*, 276(15), 11449-11452.
- Gu, Y. Z., Hogenesch, J. B., & Bradfield, C. A. (2000). The PAS superfamily: sensors of environmental and developmental signals. *Annual review of pharmacology and toxicology*, 40(1), 519-561.

- Guella, I., Sequeira, A., Rollins, B., Morgan, L., Torri, F., van Erp, T. G., ... & Akil, H. (2013). Analysis of miR-137 expression and rs1625579 in dorsolateral prefrontal cortex. *Journal of psychiatric research*, 47(9), 1215-1221.
- Guha, S., Rees, E., Darvasi, A., Ivanov, D., Ikeda, M., Bergen, S. E., ... & Cichon, S. (2013). Implication of a rare deletion at distal 16p11. 2 in schizophrenia. *JAMA psychiatry*, 70(3), 253-260.
- Guidotti, A., Auta, J., Davis, J. M., Gerevini, V. D., Dwivedi, Y., Grayson, D. R., ... & Uzunov, D. (2000). Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. *Archives of general psychiatry*, 57(11), 1061-1069.
- Gulsuner, S., Walsh, T., Watts, A. C., Lee, M. K., Thornton, A. M., Casadei, S., ... & Savage, R. M. (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*, 154(3), 518-529.
- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., ... & Wang, Y. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 464(7291), 1071-1076.
- Gur, R. E., Keshavan, M. S., & Lawrie, S. M. (2007). Deconstructing psychosis with human brain imaging. *Schizophrenia bulletin*, 33(4), 921-931.
- Gurling, H. M., Kalsi, G., Brynjolfson, J., Sigmundsson, T., Sherrington, R., Mankoo, B. S., ... & Petursson, H. (2001). Genomewide genetic linkage analysis confirms the presence of susceptibility *loci* for schizophrenia, on chromosomes 1q32. 2, 5q33. 2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23. 3-24 and 20q12. 1-11.23. *The American Journal of Human Genetics*, 68(3), 661-673.
- Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjálmsón, B. J., Xu, H., ... & Kähler, A. K. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics*, 95(5), 535-552.
- Hadland, K. A., Rushworth, M. F., Gaffan, D., & Passingham, R. E. (2003). The effect of cingulate lesions on social behaviour and emotion. *Neuropsychologia*, 41(8), 919-931.
- Häfner, H. (2003). Gender differences in schizophrenia. *Psychoneuroendocrinology*, 28, 17-54.

- Häfner, H., Löffler, W., Maurer, K., Hambrecht, M. & an der Heiden, W. (1999) Depression, negative symptoms, social stagnation and social decline in the early course of schizophrenia. *Acta Psychiatr Scand*, 100(2):105-118
- Hamada, N., Ito, H., Iwamoto, I., Morishita, R., Tabata, H., & Nagata, K. I. (2015). Role of the cytoplasmic isoform of RBFOX1/A2BP1 in establishing the architecture of the developing cerebral cortex. *Molecular autism*, 6(1), 1.
- Hamid, E. H., Hyde, T. M., Egan, M. F., Wolf, S. S., Herman, M. M., Nemeroff, C. B., & Kleinman, J. E. (2002). Neurotensin receptor binding abnormalities in the entorhinal cortex in schizophrenia and affective disorders. *Biological psychiatry*, 51(10), 795-800.
- Hamshere, M. L., Walters, J. T. R., Smith, R., Richards, A. L., Green, E., Grozeva, D., ... & Riley, B. (2013). Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Molecular psychiatry*, 18(6), 708-712.
- Hänninen, K., Katila, H., Saarela, M., Rontu, R., Mattila, K. M., Fan, M., ... & Lehtimäki, T. (2007). Interleukin-1 beta gene polymorphism and its interactions with neuregulin-1 gene polymorphism are associated with schizophrenia. *European archives of psychiatry and clinical neuroscience*, 258(1), 10-15.
- Harris, L. W., Lockstone, H. E., Khaitovich, P., Weickert, C. S., Webster, M. J., & Bahn, S. (2009). Gene expression in the prefrontal cortex during adolescence: implications for the onset of schizophrenia. *BMC medical genomics*, 2(1), 28.
- Harrison, P. J., & Weinberger, D. R. (2005). Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Molecular psychiatry*, 10(1), 40-68.
- Hass, J., Walton, E., Kirsten, H., Turner, J., Wolthusen, R., Roessner, V., ... & Ehrlich, S. (2015). Complexin2 modulates working memory-related neural activity in patients with schizophrenia. *European archives of psychiatry and clinical neuroscience*, 265(2), 137-145.
- Haukka, J., Suvisaari, J., & Lonnqvist, J. (2003). Fertility of patients with schizophrenia, their siblings, and the general population: a cohort study from 1950 to 1959 in Finland. *American Journal of Psychiatry*, 160(3), 460-463.
- Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., ... & Royall, J. J. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416), 391-399.

- Hayashi-Takagi, A., & Sawa, A. (2010). Disturbed synaptic connectivity in schizophrenia: convergence of genetic risk factors during neurodevelopment. *Brain research bulletin*, 83(3), 140-146.
- Haznedar, M. M., Buchsbaum, M. S., Hazlett, E. A., Shihabuddin, L., New, A., & Siever, L. J. (2004). Cingulate gyrus volume and metabolism in the schizophrenia spectrum. *Schizophrenia research*, 71(2), 249-262.
- Hecker, E., & Kraam, A. (2009). Hebephrenia. A contribution to clinical psychiatry'. *History of Psychiatry*, 20(2), 233-248.
- Heckers, S. Bleuler and the neurobiology of schizophrenia (2011). *Schizophr. Bull.* 37(6), 1131–1135
- Heckers, S., Barch, D. M., Bustillo, J., Gaebel, W., Gur, R., Malaspina, D., ...& Van Os, J. (2013). Structure of the psychotic disorders classification in DSM-5. *Schizophrenia research*, 150(1), 11-14.
- Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., ... & Reymond, A. (2009). Segmental copy number variation shapes tissue transcriptomes. *Nature genetics*, 41(4), 424-429.
- Herrmann, M. G., Durtschi, J. D., Bromley, L. K., Wittwer, C. T., & Voelkerding, K. V. (2006). Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. *Clinical Chemistry*, 52(3), 494-503.
- Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet*, 4(2), e1000008.
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association *loci* for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106(23), 9362-9367.
- Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6(2), 95-108.
- Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., ...& McCombie, W. R. (2007). Genome-wide in situ exon capture for selective resequencing. *Nature genetics*, 39(12), 1522-1527.

Hof, P. R., Haroutunian, V., Friedrich, V. L., Byne, W., Buitron, C., Perl, D. P., & Davis, K. L. (2003). Loss and altered spatial distribution of oligodendrocytes in the superior frontal gyrus in schizophrenia. *Biological psychiatry*, 53(12), 1075-1085.

Hoftman, G. D., Volk, D. W., Bazmi, H. H., Li, S., Sampson, A. R., & Lewis, D. A. (2015). Altered cortical expression of GABA-related genes in schizophrenia: illness progression vs developmental disturbance. *Schizophrenia bulletin*, 41(1), 180-191.

Hogenesch, J. B., Gu, Y. Z., Jain, S., & Bradfield, C. A. (1998). The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proceedings of the National Academy of Sciences*, 95(10), 5474-5479.

Horev, G., Ellegood, J., Lerch, J. P., Son, Y. E. E., Muthuswamy, L., Vogel, H., ... & Mills, A. A. (2011). Dosage-dependent phenotypes in models of 16p11. 2 lesions found in autism. *Proceedings of the National Academy of Sciences*, 108(41), 17076-17081.

Hsu, P. C., Nwulia, E., & Sawa, A. (2009). Using bioinformatic tools. *The American journal of psychiatry*, 166(8), 854.

Huang, J., Perlis, R. H., Lee, P. H., Rush, A. J., Fava, M., Sachs, G. S., ... & Purcell, S. (2010). Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *American journal of Psychiatry*.

Hubisz, M. J., & Pollard, K. S. (2014). Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Current opinion in genetics & development*, 29, 15-21.

Hudziak, J. J., Achenbach, T. M., Althoff, R. R., & Pine, D. S. (2009). A dimensional approach to developmental psychopathology. *Dimensional approaches in diagnostic classification: Refining the research agenda for DSM-V*, 101-113.

Hurles, M. E., & Lupski, J. R. (2006). Recombination hotspots in nonallelic homologous recombination. In *Genomic Disorders* (pp. 341-355). Humana Press.

Hwang, Y., Kim, J., Shin, J. Y., Kim, J. I., Seo, J. S., Webster, M. J., ... & Kim, S. (2013). Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia. *Translational psychiatry*, 3(10), e321.

Lafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., ...& Lee, C. (2004). Detection of large-scale variation in the human genome. *Nature genetics*, 36(9), 949-951.

Ikeda, M., Aleksic, B., Kinoshita, Y., Okochi, T., Kawashima, K., Kushima, I., ...& Fukuo, Y. (2011). Genome-wide association study of schizophrenia in a Japanese population. *Biological psychiatry*, 69(5), 472-478.

Illingworth, R. S., & Bird, A. P. (2009). CpG islands—‘a rough guide’. *FEBS letters*, 583(11), 1713-1720.

Impagnatiello, F., Guidotti, A. R., Pesold, C., Dwivedi, Y., Caruncho, H., Pisu, M. G., ... & Pappas, G. D. (1998). A decrease of reelin expression as a putative vulnerability factor in schizophrenia. *Proceedings of the National Academy of Sciences*, 95(26), 15718-15723.

Ingason, A., Kirov, G., Giegling, I., Hansen, T., Isles, A. R., Jakobsen, K. D., ... & Möller, H. J. (2011). Maternally derived microduplications at 15q11-q13: implication of imprinted genes in psychotic illness. *American Journal of Psychiatry*.

Ingraham, L. J., & Kety, S. S. (2000). Adoption studies of schizophrenia. *American Journal of Medical Genetics*, 97(1), 18-22.

Insel, T. R. (2010). Rethinking schizophrenia. *Nature*, 468(7321), 187-193.

International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311), 52-58.

International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299-1320.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945.

Ioannidis, J. P., Boffetta, P., Little, J., O'Brien, T. R., Uitterlinden, A. G., Vineis, P., ...& Higgins, J. P. (2008). Assessment of cumulative evidence on genetic associations: interim guidelines. *International journal of epidemiology*, 37(1), 120-132.

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., & Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. *The American Journal of Human Genetics*, 92(6), 841-853.

- Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., ...& Kendall, J. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, 74(2), 285-299.
- Irish Schizophrenia Genomics Consortium,& Wellcome Trust Case Control Consortium 2. (2012). Genome-wide association study implicates HLA-C* 01: 02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biological psychiatry*, 72(8), 620-628.
- Ishiguro, H., Koga, M., Horiuchi, Y., Noguchi, E., Morikawa, M., Suzuki, Y., ...& Inada, T. (2008). Supportive evidence for reduced expression of GNB1L in schizophrenia. *Schizophrenia bulletin*, sbn160.
- Jazbec, S., Pantelis, C., Robbins, T., Weickert, T., Weinberger, D. R., & Goldberg, T. E. (2007). Intra-dimensional/extra-dimensional set-shifting performance in schizophrenia: impact of distractors. *Schizophrenia research*, 89(1), 339-349.
- Ji, J., Sundquist, K., Ning, Y., Kendler, K. S., Sundquist, J., & Chen, X. (2013). Incidence of cancer in patients with schizophrenia and their first-degree relatives: a population-based study in Sweden. *Schizophrenia bulletin*, 39(3), 527-536.
- Johannes, F., Colot, V., & Jansen, R. C. (2008). Epigenome dynamics: a quantitative genetics perspective. *Nature Reviews Genetics*, 9(11), 883-890.
- Johnson, G. C., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., Di Genova, G., ... & Twells, R. C. (2001). Haplotype tagging for the identification of common disease genes. *Nature genetics*, 29(2), 233-237.
- Johnson, M. B., Kawasawa, Y. I., Mason, C. E., Krsnik, Ž., Coppola, G., Bogdanović, D., ... & Šestan, N. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*, 62(4), 494-509.
- Johnson, R. D., Oliver, P. L., & Davies, K. E. (2008). SNARE proteins and schizophrenia: linking synaptic and neurodevelopmental hypotheses. *Acta biochimica Polonica*, 55(4), 619-628.
- Jomphe, C., Lemelin, P. L., Okano, H., Kobayashi, K., & Trudeau, L. E. (2006). Bidirectional regulation of dopamine D2 and neurotensin NTS1 receptors in dopamine neurons. *European Journal of Neuroscience*, 24(10), 2789-2800.
- Juhasz, G., Dunham, J. S., McKie, S., Thomas, E., Downey, D., Chase, D., ... & Payton, A. (2011). The CREB1-BDNF-NTRK2 pathway in depression: multiple gene-cognition-environment interactions. *Biological psychiatry*, 69(8), 762-771.

- Juraeva, D., Haenisch, B., Zapatka, M., Frank, J., Witt, S. H., Mühleisen, T. W., ... & Giegling, I. (2014). Integrated pathway-based approach identifies association between genomic regions at CTCF and CACNB2 and schizophrenia. *PLoS Genet*, 10(6), e1004345.
- Kamm, G. B., López-Leal, R., Lorenzo, J. R., & Franchini, L. F. (2013b). A fast-evolving human NPAS3 enhancer gained reporter expression in the developing forebrain of transgenic mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1632), 20130019.
- Kamm, G. B., Pisciotto, F., Kliger, R., & Franchini, L. F. (2013a). The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Molecular biology and evolution*, mst023.
- Kamnasaran, D., Muir, W. J., Ferguson-Smith, M. A., & Cox, D. W. (2003). Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. *Journal of medical genetics*, 40(5), 325-332.
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), 27-30.
- Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., ... & Guennel, T. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, 478(7370), 483-489.
- Kano, S., Colantuoni, C., Han, F., Zhou, Z., Yuan, Q., Wilson, A., ... & Cascella, N. (2013). Genome-wide profiling of multiple histone methylations in olfactory cells: further implications for cellular susceptibility to oxidative stress in schizophrenia. *Molecular psychiatry*, 18(7), 740.
- Kapur, S., & Remington, G. (1996). Serotonin-dopamine interaction and its relevance to schizophrenia. *American Journal of Psychiatry*, 153(4), 466-476.
- Karayorgou, M., Morris, M. A., Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J., ... & Lasseter, V. K. (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proceedings of the National Academy of Sciences*, 92(17), 7612-7616.
- Karlsson, H., Bachmann, S., Schröder, J., McArthur, J., Torrey, E. F., & Yolken, R. H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proceedings of the National Academy of Sciences*, 98(8), 4634-4639.
- Kasanin, J. (1933). The acute schizoaffective psychoses. *American Journal of Psychiatry*, 90(1), 97-126.

Bibliografía

Katsel, P., Byne, W., Roussos, P., Tan, W., Siever, L., & Haroutunian, V. (2011). Astrocyte and glutamate markers in the superficial, deep, and white matter layers of the anterior cingulate gyrus in schizophrenia. *Neuropsychopharmacology*, 36(6), 1171-1177.

Katsel, P., Davis, K. L., & Haroutunian, V. (2005). Variations in myelin and oligodendrocyte-related gene expression across multiple brain regions in schizophrenia: a gene ontology study. *Schizophrenia research*, 79(2), 157-173.

Kaur, H., Jajodia, A., Grover, S., Baghel, R., Gupta, M., Jain, S., & Kukreti, R. (2014). Genetic variations of PIP4K2A confer vulnerability to poor antipsychotic response in severely ill schizophrenia patients. *PloS one*, 9(7), e102556.

Kavanagh, D. H., Tansey, K. E., O'Donovan, M. C., & Owen, M. J. (2014). Schizophrenia genetics: emerging themes for a complex disorder. *Molecular psychiatry*.

Kendler, K. S., Myers, J. M., O'Neill, F. A., Martin, R., Murphy, B., MacLean, C. J., ... & Straub, R. E. (2000). Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. *American Journal of Psychiatry*.

Kerr, M. K., & Churchill, G. A. (2001). Statistical design and the analysis of gene expression microarray data. *Genetical research*, 77(02), 123-128.

Keshavan, M. S., Nasrallah, H. A., & Tandon, R. (2011). Schizophrenia, "Just the Facts" 6. Moving ahead with the schizophrenia concept: from the elephant to the mouse. *Schizophrenia research*, 127(1), 3-13.

Kilpinen, H., Waszak, S. M., Gschwind, A. R., Raghav, S. K., Witwicki, R. M., Orioli, A., ... & Yurovsky, A. (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science*, 342(6159), 744-747.

Kim, A. H., Parker, E. K., Williamson, V., McMichael, G. O., Fanous, A. H., & Vladimirov, V. I. (2012). Experimental validation of candidate schizophrenia gene ZNF804A as target for hsa-miR-137. *Schizophrenia research*, 141(1), 60-64.

Kim, C. W., Tak, H. M., Kim, G. T., Mun, Y. J., Jeon, B. T., Kim, H. J., ... & Kang, D. (2010). H2O2-induced up-regulation of CatSper3 in mouse brain. *Molecular reproduction and development*, 77(8), 650-650.

- Kim, J. S., Kornhuber, H. H., Schmid-Burgk, W., & Holzmüller, B. (1980). Low cerebrospinal fluid glutamate in schizophrenic patients and a new hypothesis on schizophrenia. *Neuroscience letters*, 20(3), 379-382.
- King, M. C., & Wilson, A. C. (1975). *Evolution at two levels in humans and chimpanzees*. na.
- Kirkpatrick, B., & Miller, B. J. (2013). Inflammation and schizophrenia. *Schizophrenia bulletin*, 39(6), 1174-1179.
- Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., ...& Wellcome Trust Case Control Consortium. (2009a). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Human molecular genetics*, 18(8), 1497-1503.
- Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., ...& Grozeva, D. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular psychiatry*, 17(2), 142-153.
- Kirov, G., Zaharieva, I., Georgieva, L., Moskvina, V., Nikolov, I., Cichon, S., ...& O'Donovan, M. C. (2009b). A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Molecular psychiatry*, 14(8), 796-803.
- Kitabayashi, M., Nishiya, Y., Esaka, M., Itakura, M., & Imanaka, T. (2002). Gene cloning and polymerase chain reaction with proliferating cell nuclear antigen from *Thermococcus kodakaraensis* KOD1. *Bioscience, biotechnology, and biochemistry*, 66(10), 2194-2200.
- Kleer, C. G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S. A., ... & Sabel, M. S. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proceedings of the National Academy of Sciences*, 100(20), 11606-11611.
- Knable, M. B., Barci, B. M., Webster, M. J., Meador-Woodruff, J., & Torrey, E. F. (2004). Molecular abnormalities of the hippocampus in severe psychiatric illness: postmortem findings from the Stanley Neuropathology Consortium. *Molecular psychiatry*, 9(6), 609-620.
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., ... & Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*, 22(3), 568-576.
- Kobrynski, L. J., & Sullivan, K. E. (2007). Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *The Lancet*, 370(9596), 1443-1452.

Kocerha, J., Kauppinen, S., & Wahlestedt, C. (2009). microRNAs in CNS disorders. *Neuromolecular medicine*, 11(3), 162-172.

Kochunov, P., Glahn, D. C., Rowland, L. M., Olvera, R. L., Winkler, A., Yang, Y. H., ... & Blangero, J. (2013). Testing the hypothesis of accelerated cerebral white matter aging in schizophrenia and major depression. *Biological psychiatry*, 73(5), 482-491.

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., ...& Wong, W. S. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488(7412), 471-475.

Konopaske, G. T., Dorph-Petersen, K. A., Sweet, R. A., Pierri, J. N., Zhang, W., Sampson, A. R., & Lewis, D. A. (2008). Effect of chronic antipsychotic exposure on astrocyte and oligodendrocyte numbers in macaque monkeys. *Biological psychiatry*, 63(8), 759-765.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4), 693-705.

Kraepelin E (1899). *Psychiatrie. Ein Lehrbuch für Studierende und Ärzte*, 6. Auflage, 1. Band. Barth : Leipzig. [trans. H. Metoui: *Psychiatry, A Textbook for Students and Physicians*, 6th edn, Vol. 1. Science History Publications : Canton, MA, 1990].

Kruglyak, L. (1999). Genetic isolates: Separate but equal?. *Proceedings of the National Academy of Sciences*, 96(4), 1170-1172.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., ... & Adli, M. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet*, 4(10), e1000242.

Kumarasinghe, N., Tooney, P. A., & Schall, U. (2012). Finding the needle in the haystack: A review of microarray gene expression research into schizophrenia. *Australian and New Zealand Journal of Psychiatry*, 46(7), 598-610.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., ...& Amin, V. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539), 317-330.

Kwok, P. Y. (2001). Methods for genotyping single nucleotide polymorphisms. *Annual review of genomics and human genetics*, 2(1), 235-258.

Kwon, E., Wang, W., & Tsai, L. H. (2013). Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Molecular psychiatry*, 18(1), 11-12.

- Lahti, A. C., Weiler, M. A., Tamara, M., Parwani, A., & Tamminga, C. A. (2001). Effects of ketamine in normal and schizophrenic volunteers. *Neuropsychopharmacology*, 25(4), 455-467.
- Lambert, M., Karow, A., Leucht, S., Schimmelmann, B. G., & Naber, D. (2010). Remission in schizophrenia: validity, frequency, predictors, and patients' perspective 5 years later. *Dialogues in clinical neuroscience*, 12(3), 393.
- Lander, E. S. (1996). The new genomics: global views of biology. *Science*, 274(5287), 536-539.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... & Funke, R. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921.
- Langfeldt, G. (1939). The schizophrenic states. *Munksgaard, Copenhagen*.
- Lanzuolo, C., & Orlando, V. (2012). Memories from the polycomb group proteins. *Annual review of genetics*, 46, 561-589.
- Lee, H. J., Song, J. Y., Kim, J. W., Jin, S. Y., Hong, M. S., Park, J. K., ... & Fukumaki, Y. (2005). Association study of polymorphisms in synaptic vesicle-associated genes, SYN2 and CPLX2, with schizophrenia. *Behavioral and Brain Functions*, 1(1), 1.
- Lee, J. A., Carvalho, C. M., & Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *cell*, 131(7), 1235-1247.
- Lee, J. A., Damianov, A., Lin, C. H., Fontes, M., Parikshak, N. N., Anderson, E. S., ... & Martin, K. C. (2016). Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron*, 89(1), 113-128.
- Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., ... & NHLBI GO Exome Sequencing Project. (2012a). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *The American Journal of Human Genetics*, 91(2), 224-237.
- Lee, Y., Mattai, A., Long, R., Rapoport, J. L., Gogtay, N., & Addington, A. M. (2012b). Microduplications disrupting the MYT1L gene (2p25.3) are associated with schizophrenia. *Psychiatric genetics*, 22(4), 206.
- Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., ... & Tukiainen, T. (2015). Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv*, 030338.

Bibliografía

- Lencz, T., Guha, S., Liu, C., Rosenfeld, J., Mukherjee, S., DeRosse, P., ...& Ikeda, M. (2013). Genome-wide association study implicates NDST3 in schizophrenia and bipolar disorder. *Nature communications*, 4.
- Lencz, T., Morgan, T. V., Athanasiou, M., Dain, B., Reed, C. R., Kane, J. M., ...& Malhotra, A. K. (2007). Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Molecular psychiatry*, 12(6), 572-580.
- Lennertz, L., Quednow, B. B., Benninghoff, J., Wagner, M., Maier, W., & Mössner, R. (2011). Impact of TCF4 on the genetics of schizophrenia. *European archives of psychiatry and clinical neuroscience*, 261(2), 161-165.
- Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neuroscience & Biobehavioral Reviews*, 30(6), 718-729.
- Leung, M. D., & Psych, C. M. (2000). Sex differences in schizophrenia, a review of the literature. *Acta Psychiatrica Scandinavica*, 101(401), 3-38.
- Levinson, D. F., Duan, J., Oh, S., Wang, K., Sanders, A. R., Shi, J., ...& Cloninger, C. R. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *American Journal of Psychiatry*.
- Lewis, D. A., & Hashimoto, T. (2007). Deciphering the disease process of schizophrenia: the contribution of cortical GABA neurons. *International review of neurobiology*, 78, 109-131.
- Lewis, D. A., & Levitt, P. (2002). Schizophrenia as a disorder of neurodevelopment. *Annual review of Neuroscience*, 25(1), 409-432.
- Lewis, D. A., Hashimoto, T., & Volk, D. W. (2005). Cortical inhibitory neurons and schizophrenia. *Nature Reviews Neuroscience*, 6(4), 312-324.
- Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. In *Genes, Development and Cancer* (pp. 205-217). Springer US.
- Lewis, P. H. (1949). Pc: Polycomb. *Drosophila Information Service*, 21, 69.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Li, J., Chen, C., Chen, C., He, Q., Li, H., Li, J., ... & Dong, Q. (2011). Neurotensin receptor 1 gene (NTSR1) polymorphism is associated with working memory. *PLoS One*, 6(3), e17365.

- Li, J., Chen, C., Lei, X., Wang, Y., Chen, C., He, Q., ... & Dong, Q. (2013). The NTSR1 gene modulates the association between hippocampal structure and working memory performance. *Neuroimage*, 75, 79-86.
- Li, J., You, Y., Yue, W., Yu, H., Lu, T., Wu, Z., ... & Wang, L. (2016a). Chromatin remodeling gene EZH2 involved in the genetic etiology of autism in Chinese Han population. *Neuroscience letters*, 610, 182-186.
- Li, W. X., Dai, S. X., Liu, J. Q., Wang, Q., Li, G. H., & Huang, J. F. (2016b). Integrated Analysis of Alzheimer's Disease and Schizophrenia Dataset Revealed Different Expression Pattern in Learning and Memory. *Journal of Alzheimer's Disease*, (Preprint), 1-9.
- Liao, B. Y., & Zhang, J. (2008). Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proceedings of the National Academy of Sciences*, 105(19), 6987-6992.
- Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*, 373(9659), 234-239.
- Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end joining pathway. *Annual review of biochemistry*, 79, 181.
- Lin, C. Y., Li, M. J., Chang, J. G., Liu, S. C., Weng, T., Wu, K. H., ... & Peng, C. T. (2014). High-resolution melting analyses for genetic variants in ARID5B and IKZF1 with childhood acute lymphoblastic leukemia susceptibility loci in Taiwan. *Blood Cells, Molecules, and Diseases*, 52(2), 140-145.
- Lin, G. M., Chen, Y. J., Kuo, D. J., Jaiteh, L. E., Wu, Y. C., Lo, T. S., & Li, Y. H. (2013a). Cancer incidence in patients with schizophrenia or bipolar disorder: a nationwide population-based study in Taiwan, 1997–2009. *Schizophrenia bulletin*, 39(2), 407-416.
- Lin, Z., Su, Y., Zhang, C., Xing, M., Ding, W., Liao, L., ... & Cui, D. (2013b). The interaction of BDNF and NTRK2 gene increases the susceptibility of paranoid schizophrenia. *PloS one*, 8(9), e74264.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., ... & Ward, L. D. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478(7370), 476-482.

- Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., ... & Mauceli, E. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069), 803-819.
- Linderholm, K. R., Skogh, E., Olsson, S. K., Dahl, M. L., Holtze, M., Engberg, G., ... & Erhardt, S. (2010). Increased levels of kynurenine and kynurenic acid in the CSF of patients with schizophrenia. *Schizophrenia bulletin*, sbq086.
- Linscott, R. J., Allardyce, J., & van Os, J. (2010). Seeking verisimilitude in a class: a systematic review of evidence that the criterial clinical symptoms of schizophrenia are taxonic. *Schizophrenia bulletin*, 36(4), 811-829.
- Liu, J., Lewohl, J. M., Harris, R. A., Iyer, V. R., Dodd, P. R., Randall, P. K., & Mayfield, R. D. (2006a). Patterns of gene expression in the frontal cortex discriminate alcoholic from nonalcoholic individuals. *Neuropsychopharmacology*, 31(7), 1574-1582.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., ... & Law, M. (2012). Comparison of next-generation sequencing systems. *BioMed Research International*, 2012.
- Liu, S., Rauhut, R., Vornlocher, H. P., & Lührmann, R. (2006b). The network of protein-protein interactions within the human U4/U6. U5 tri-snRNP. *Rna*, 12(7), 1418-1430.
- Liu, Z., Li, X., Sun, N., Xu, Y., Meng, Y., Yang, C., ... & Zhang, K. (2014). Microarray profiling and co-expression network analysis of circulating lncRNAs and mRNAs associated with major depressive disorder. *PloS one*, 9(3), e93388.
- Lo, W. S., Xu, Z., Yu, Z., Pun, F. W., Ng, S. K., Chen, J., ... & Harano, M. (2007). Positive selection within the Schizophrenia-associated GABA (A) receptor beta2 gene. *PLoS One*, 2(5), e462-e462.
- Lokmane, L., Heliot, C., Garcia-Villalba, P., Fabre, M., & Cereghini, S. (2010). vHNF1 functions in distinct regulatory circuits to control ureteric bud branching and early nephrogenesis. *Development*, 137(2), 347-357.
- Long, F., Wang, X., & Chen, X. (2013). Mutation screening of the DISC1 exonic sequence in 96 schizophrenics and 96 healthy controls using high-resolution melting analysis. *Journal of Psychiatry*, 1, 006.
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., ... & Foster, B. (2013). The genotype-tissue expression (GTEx) project. *Nature genetics*, 45(6), 580-585.

- Lundgaard, I., Luzhynskaya, A., Stockley, J. H., Wang, Z., Evans, K. A., Swire, M., ... & Káradóttir, R. T. (2013). Neuregulin and BDNF induce a switch to NMDA receptor-dependent myelination by oligodendrocytes. *PLoS Biol*, 11(12), e1001743.
- Lunter, G., Ponting, C. P., & Hein, J. (2006). Genome-wide identification of human functional DNA using a neutral indel model. *PLoS Comput Biol*, 2(1), e5.
- Luo, X. J., Huang, L., van den Oord, E. J., Aberg, K. A., Gan, L., Zhao, Z., & Yao, Y. G. (2014). Common variants in the MKL1 gene confer risk of schizophrenia. *Schizophrenia bulletin*, sbu156.
- Lupski, J. R. (2010). Retrotransposition and structural variation in the human genome. *Cell*, 141(7), 1110-1112.
- Lykouras, L., Oulis, P., Daskalopoulou, E., Psarros, K., & Christodoulou, G. N. (2001). Clinical subtypes of schizophrenic disorders: a cluster analytic study. *Psychopathology*, 34(1), 23-28.
- Lynall, M. E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., & Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. *The Journal of Neuroscience*, 30(28), 9477-9487.
- Lynch, M. D., Smith, A. J., De Gobbi, M., Flenley, M., Hughes, J. R., Vernimmen, D., ... & Meek, S. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO journal*, 31(2), 317-329.
- Ma, H., Huang, Y., Zhang, B., Li, J., Wang, Y., Zhao, X., ... & Zhu, G. (2013). Association between neurotensin receptor 1 (NTR1) gene polymorphisms and schizophrenia in a Han Chinese population. *Journal of Molecular Neuroscience*, 50(2), 345-352.
- Ma, X., Deng, W., Liu, X., Li, M., Chen, Z., He, Z., ... & Li, T. (2011). A genome-wide association study for quantitative traits in schizophrenia in China. *Genes, Brain and Behavior*, 10(7), 734-739.
- MacIntyre, D. J., Blackwood, D. H. R., Porteous, D. J., Pickard, B. S., & Muir, W. J. (2003). Chromosomal abnormalities and mental illness. *Molecular psychiatry*, 8(3), 275-287.
- Macintyre, G., Alford, T., Xiong, L., Rouleau, G. A., Tibbo, P. G., & Cox, D. W. (2010). Association of NPAS3 exonic variation with schizophrenia. *Schizophrenia research*, 120(1), 143-149.
- Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature News*, 456(7218), 18-21.

Bibliografía

- Mai, J. K., Triepel, J., & Metz, J. (1987). Neurotensin in the human brain. *Neuroscience*, 22(2), 499-524.
- Makari GJ. Revolution in Mind: The Creation of Psychoanalysis. New York, NY: Harper Collins; 2008.
- Makino, Y., Cao, R., Svensson, K., Bertilsson, G., Asman, M., Tanaka, H., ... & Poellinger, L. (2001). Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression. *Nature*, 414(6863), 550-554.
- Malaspina, D., Gilman, C., & Kranz, T. M. (2015). Paternal age and mental health of offspring. *Fertility and sterility*.
- Malhotra, D., & Sebat, J. (2012). CNV: harbingers of a rare variant revolution in psychiatric genetics. *Cell*, 148(6), 1223-1241.
- Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., ... & Gill, M. (2011). High frequencies of de novo CNV in bipolar disorder and schizophrenia. *Neuron*, 72(6), 951-963.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Cho, J. H. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747-753.
- Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in genetics*, 24(3), 133-141.
- Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, 469(7330), 343-349.
- Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., ... & Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular cell*, 32(4), 503-518.
- Marta, J. (1987). Lacan and post-Structuralism. *The American journal of psychoanalysis*, 47(1), 51-57.
- Martin, C. L., Duvall, J. A., Ilkin, Y., Simon, J. S., Arreaza, M. G., Wilkes, K., ... & Cook, E. (2007). Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 144(7), 869-876.

- Martin, M. V., Mirnics, K., Nisenbaum, L. K., & Vawter, M. P. (2015). Olanzapine reversed brain gene expression changes induced by phencyclidine treatment in non-human primates. *Molecular neuropsychiatry*, 1(2), 82-93.
- Martorell, L., Costas, J., Valero, J., Gutierrez-Zotes, A., Phillips, C., Torres, M., ... & Vallès, V. (2008). Analyses of variants located in estrogen metabolism genes (ESR1, ESR2, COMT and APOE) and schizophrenia. *Schizophrenia research*, 100(1), 308-315.
- Matute, C., Melone, M., Vallejo-Illarramendi, A., & Conti, F. (2005). Increased expression of the astrocytic glutamate transporter GLT-1 in the prefrontal cortex of schizophrenics. *Glia*, 49(3), 451-455.
- Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., ... & Shafer, A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337(6099), 1190-1195.
- Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564.
- Maycox, P. R., Kelly, F., Taylor, A., Bates, S., Reid, J., Logendra, R., ... & Davies, C. (2009). Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Molecular psychiatry*, 14(12), 1083-1094.
- McCarthy, S. E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Bernstein, Y., ... & Antoniou, E. (2014). De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular psychiatry*, 19(6), 652-658.
- McClellan, J. M., Susser, E., & King, M. C. (2007). Schizophrenia: a common disease caused by multiple rare alleles. *The British Journal of Psychiatry*, 190(3), 194-199.
- McCullumsmith, R. E., O'Donovan, S. M., Drummond, J. B., Benesh, F. S., Simmons, M., Roberts, R., ... & Meador-Woodruff, J. H. (2015). Cell-specific abnormalities of glutamate transporters in schizophrenia: sick astrocytes and compensating relay neurons & quest. *Molecular psychiatry*.
- McGlashan, T. H., & Johannessen, J. O. (1996). Early detection and intervention with schizophrenia: rationale. *Schizophrenia bulletin*, 22(2), 201-222.
- McGrath, J. J. (2005). Myths and plain truths about schizophrenia epidemiology—the NAPE lecture 2004. *Acta Psychiatrica Scandinavica*, 111(1), 4-11.

Bibliografía

- McGrath, J. J., & Sasser, E. S. (2009). New directions in the epidemiology of schizophrenia. *Medical Journal of Australia*, 190(4), S7.
- McGrath, J. J., Petersen, L., Agerbo, E., Mors, O., Mortensen, P. B., & Pedersen, C. B. (2014). A comprehensive assessment of parental age and psychiatric disorders. *JAMA psychiatry*, 71(3), 301-309.
- McGrath, J., Eyles, D., Mowry, B., Yolken, R., & Buka, S. (2003). Low maternal vitamin D as a risk factor for schizophrenia: a pilot study using banked sera. *Schizophrenia research*, 63(1), 73-78.
- McInerney, P., Adams, P., & Hadi, M. Z. (2014). Error rate comparison during polymerase chain reaction by DNA polymerase. *Molecular biology international*, 2014.
- McVicker, G., van de Geijn, B., Degner, J. F., Cain, C. E., Banovich, N. E., Raj, A., ... & Pritchard, J. K. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science*, 342(6159), 747-749.
- Meador, S., Ponting, C. P., & Lunter, G. (2010). Massive turnover of functional sequence in human and other mammalian genomes. *Genome research*, 20(10), 1335-1343.
- Mehta, D., Iwamoto, K., Ueda, J., Bundo, M., Adati, N., Kojima, T., & Kato, T. (2014). Comprehensive survey of CNV influencing gene expression in the human brain and its implications for pathophysiology. *Neuroscience research*, 79, 22-33.
- Melzer, D., Perry, J. R., Hernandez, D., Corsi, A. M., Stevens, K., Rafferty, I., ... & Rafiq, S. (2008). A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet*, 4(5), e1000072.
- Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., ... & Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet*, 6(12), e1001244.
- Messias, E. L., Chen, C. Y., & Eaton, W. W. (2007). Epidemiology of schizophrenia: review of findings and myths. *Psychiatric Clinics of North America*, 30(3), 323-338.
- Meyer, U., Feldon, J., & Yee, B. K. (2009). A review of the fetal brain cytokine imbalance hypothesis of schizophrenia. *Schizophrenia bulletin*, 35(5), 959-972.
- Meyer, U., Feldon, J., & Dammann, O. (2011). Schizophrenia and autism: both shared and disorder-specific pathogenesis via perinatal inflammation?. *Pediatric research*, 69, 26R-33R.

- Meyer-Lindenberg, A. (2010). From maps to mechanisms through neuroimaging of schizophrenia. *Nature*, 468(7321), 194-202.
- Migliavacca, E., Golzio, C., Männik, K., Blumenthal, I., Oh, E. C., Harewood, L., ... & Maillard, A. M. (2015). A potential contributory role for ciliary dysfunction in the 16p11. 2 600 kb BP4-BP5 pathology. *The American Journal of Human Genetics*, 96(5), 784-796.
- Mikhail, F. M., Lose, E. J., Robin, N. H., Descartes, M. D., Rutledge, K. D., Rutledge, S. L., ... & Carroll, A. J. (2011). Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. *American Journal of medical Genetics Part A*, 155(10), 2386-2396.
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., ... & Lee, W. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153), 553-560.
- Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., ... & Wang, S. C. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *The American Journal of Human Genetics*, 82(3), 696-711.
- Millar, J. K., Wilson-Annan, J. C., Anderson, S., Christie, S., Taylor, M. S., Semple, C. A., ... & Porteous, D. J. (2000). Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Human molecular genetics*, 9(9), 1415-1423.
- Miller, B. J., Buckley, P., Seabolt, W., Mellor, A., & Kirkpatrick, B. (2011). Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biological psychiatry*, 70(7), 663-671.
- Miller, B. J., Culpepper, N., Rapaport, M. H., & Buckley, P. (2013a). Prenatal inflammation and neurodevelopment in schizophrenia: a review of human studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 42, 92-100.
- Miller, B. J., Gassama, B., Sebastian, D., Buckley, P., & Mellor, A. (2013b). Meta-analysis of lymphocytes in schizophrenia: clinical status and antipsychotic effects. *Biological psychiatry*, 73(10), 993-999.
- Miller, J. A., Ding, S. L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., ... & Pletikos, M. (2014). Transcriptional landscape of the prenatal human brain. *Nature*, 508(7495), 199-206.

Bibliografía

- Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., ... & Chinwalla, A. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, 470(7332), 59-65.
- Mirnics, K., Middleton, F. A., Marquez, A., Lewis, D. A., & Levitt, P. (2000). Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron*, 28(1), 53-67.
- Missler, M., Zhang, W., Rohlmann, A., Kattenstroth, G., Hammer, R. E., Gottmann, K., & Südhof, T. C. (2003). α -Neurexins couple Ca^{2+} channels to synaptic vesicle exocytosis. *Nature*, 423(6943), 939-948.
- Mistry, M., Gillis, J., & Pavlidis, P. (2013a). Genome-wide expression profiling of schizophrenia using a large combined cohort. *Molecular psychiatry*, 18(2), 215-225.
- Mistry, M., Gillis, J., & Pavlidis, P. (2013b). Meta-analysis of gene coexpression networks in the post-mortem prefrontal cortex of patients with schizophrenia and unaffected controls. *BMC neuroscience*, 14(1), 105.
- Montes, J.M., y Saiz, J. Esquizofrenia. En: Barcia D. Dir. Tratado dePsiquiatría. Madrid: Arán Editores; 2000.
- Moreno-De-Luca, D., Sanders, S. J., Willsey, A. J., Mulle, J. G., Lowe, J. K., Geschwind, D. H., ... & Ledbetter, D. H. (2013). Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. *Molecular psychiatry*, 18(10), 1090-1095.
- Morgan, V. A., Castle, D. J., & Jablensky, A. V. (2008). Do women express and experience psychosis differently from men? Epidemiological evidence from the Australian National Study of Low Prevalence (Psychotic) Disorders. *Australian and New Zealand Journal of Psychiatry*, 42(1), 74-82.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), 621-628.
- Mortensen, P. B., Nørgaard-Pedersen, B., Waltoft, B. L., Sørensen, T. L., Hougaard, D., & Yolken, R. H. (2007). Early infections of *Toxoplasma gondii* and the later development of schizophrenia. *Schizophrenia bulletin*, 33(3), 741-744.
- Mulle, J. G., Dodd, A. F., McGrath, J. A., Wolyniec, P. S., Mitchell, A. A., Shetty, A. C., ... & Cutler, D. J. (2010). Microdeletions of 3q29 confer high risk for schizophrenia. *The American Journal of Human Genetics*, 87(2), 229-236.

- Müller, N., & Schwarz, M. J. (2008). A psychoneuroimmunological perspective to Emil Kraepelin's dichotomy. *European Archives of Psychiatry and clinical neuroscience*, 258, 97-106.
- Murphy, K. C., Jones, L. A., & Owen, M. J. (1999). High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Archives of general psychiatry*, 56(10), 940-945.
- Murray, R. M., & Lewis, S. W. (1987). Is schizophrenia a neurodevelopmental disorder?. *BMJ*, 295(6600), 681-682.
- Murray, R. M., Morrison, P. D., Henquet, C., & Di Forti, M. (2007). Cannabis, the mind and society: the hash realities. *Nature Reviews Neuroscience*, 8(11), 885-895.
- Mutsuddi, M., Morris, D. W., Waggoner, S. G., Daly, M. J., Scolnick, E. M., & Sklar, P. (2006). Analysis of high-resolution HapMap of DTNBP1 (Dysbindin) suggests no consistency between reported common variant associations and schizophrenia. *The American Journal of Human Genetics*, 79(5), 903-909.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), 1344-1349.
- Najas-García, A., Rufián, S., & Rojo, E. (2014). Neurodevelopment or neurodegeneration: review of theories of schizophrenia. *Actas Esp Psiquiatr*, 42(4), 185-95.
- Nakatani, J., Tamada, K., Hatanaka, F., Ise, S., Ohta, H., Inoue, K., ... & Iwamoto, K. (2009). Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. *Cell*, 137(7), 1235-1246.
- Neale, B. M., Medland, S. E., Ripke, S., Asherson, P., Franke, B., Lesch, K. P., ... & Daly, M. (2010). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(9), 884-897.
- Need, A. C., Ge, D., Weale, M. E., Maia, J., Feng, S., Heinzen, E. L., ... & Strittmatter, W. J. (2009). A genome-wide investigation of SNP and CNV in schizophrenia. *PLoS Genet*, 5(2), e1000373.
- Need, A. C., McEvoy, J. P., Gennarelli, M., Heinzen, E. L., Ge, D., Maia, J. M., ... & Zhao, Q. (2012). Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *The American Journal of Human Genetics*, 91(2), 303-312.

Bibliografía

- Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., ... & Maurano, M. T. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, 489(7414), 83-90.
- Network, T., & Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. (2015). Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature neuroscience*, 18(2), 199-209.
- Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, 31(13), 3812-3814.
- Nguyen, D. Q., Webber, C. P., Hehir-Kwa, J., Pfundt, R., Veltman, J., & Ponting, C. P. (2008). Reduced purifying selection prevails over positive selection in human copy number variant evolution. *Genome research*, gr-077289.
- Nikkilä, H. V., Müller, K., Ahokas, A., Rimón, R., & Andersson, L. C. (2001). Increased frequency of activated lymphocytes in the cerebrospinal fluid of patients with acute schizophrenia. *Schizophrenia research*, 49(1), 99-105.
- Nishimura, D. (2001). BioCarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient*, 2(3), 117-120.
- Nord, A. S., Pattabiraman, K., Visel, A., & Rubenstein, J. L. (2015). Genomic Perspectives of Transcriptional Regulation in Forebrain Development. *Neuron*, 85(1), 27-47.
- O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., ... & Dwyer, S. (2008). Identification of *loci* associated with schizophrenia by genome-wide association and follow-up. *Nature genetics*, 40(9), 1053-1055.
- O'Donovan, M. C., Craddock, N. J., & Owen, M. J. (2009). Genetics of psychosis; insights from views across the genome. *Human genetics*, 126(1), 3-12.
- O'Connor, J. A., Muly, E. C., Arnold, S. E., & Hemby, S. E. (2007). AMPA receptor subunit and splice variant expression in the DLPFC of schizophrenic subjects and rhesus monkeys chronically administered antipsychotic drugs. *Schizophrenia research*, 90(1), 28-40.
- Oh, S. K., Baek, J. I., Weigand, K. M., Venselaar, H., Swarts, H. G., Park, S. H., ... & Vriend, G. (2015). A missense variant of the ATP1A2 gene is associated with a novel phenotype of progressive sensorineural hearing loss associated with migraine. *European Journal of Human Genetics*, 23(5), 639-645.

- Okoniewski, M. J., & Miller, C. J. (2006). Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC bioinformatics*, 7(1), 276.
- Oldham, M. C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., & Geschwind, D. H. (2008). Functional organization of the transcriptome in human brain. *Nature neuroscience*, 11(11), 1271-1282.
- Oleksiak, M. F., Churchill, G. A., & Crawford, D. L. (2002). Variation in gene expression within and among natural populations. *Nature genetics*, 32(2), 261-266.
- Olney, J. W., & Farber, N. B. (1995). Glutamate receptor dysfunction and schizophrenia. *Archives of general psychiatry*, 52(12), 998-1007.
- Ordemann, G. J., Oppen, J., & Davalos, D. (2014). Prospective memory in schizophrenia: a review. *Schizophrenia research*, 155(1), 77-89.
- Orellana, G., & Slachevsky, A. (2013). Executive functioning in schizophrenia. *Frontiers in psychiatry*, 4.
- Ota, V. K., Gadelha, A., Assunção, I. B., Santoro, M. L., Christofolini, D. M., Bellucco, F. T., ...& Melaragno, M. I. (2013). ZDHHC8 gene may play a role in cortical volumes of patients with schizophrenia. *Schizophrenia research*, 145(1), 33-35.
- Ouchi, Y., Banno, Y., Shimizu, Y., Ando, S., Hasegawa, H., Adachi, K., & Iwamoto, T. (2013). Reduced adult hippocampal neurogenesis and working memory deficits in the Dgcr8-deficient mouse model of 22q11. 2 deletion-associated schizophrenia can be rescued by IGF2. *The Journal of Neuroscience*, 33(22), 9408-9419.
- Owen, M. J., Craddock, N., & O'Donovan, M. C. (2005a). Schizophrenia: genes at last?. *TRENDS in Genetics*, 21(9), 518-525.
- Owen, M. J., O'Donovan, M. C., Thapar, A., & Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. *The British Journal of Psychiatry*, 198(3), 173-175.
- Owen, M. J., O'Donovan, M. C., & Harrison, P. J. (2005b). Schizophrenia: a genetic disorder of the synapse?. *Bmj*, 330(7484), 158-159.
- Papiol, S., Molina, V., Desco, M., Rosa, A., Reig, S., Gispert, J. D., ...& Fañanás, L. (2005). Ventricular enlargement in schizophrenia is associated with a genetic polymorphism at the interleukin-1 receptor antagonist gene. *Neuroimage*, 27(4), 1002-1006.

- Park, C., Qian, W., & Zhang, J. (2012). Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO reports*, 13(12), 1123-1129.
- Parker, S. C., Hansen, L., Abaan, H. O., Tullius, T. D., & Margulies, E. H. (2009). Local DNA topography correlates with functional noncoding regions of the human genome. *Science*, 324(5925), 389-392.
- Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., ... & Nguyen, B. T. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, 294(5547), 1719-1723.
- Patterson, P. H. (2009). Immune involvement in schizophrenia and autism: etiology, pathology and animal models. *Behavioural brain research*, 204(2), 313-321.
- Paul-Samojedny, M., Kowalczyk, M., Suchanek, R., Owczarek, A., Fila-Danilow, A., Szczygiel, A., & Kowalski, J. (2010). Functional polymorphism in the interleukin-6 and interleukin-10 genes in patients with paranoid schizophrenia—a case-control study. *Journal of molecular neuroscience*, 42(1), 112-119.
- Peippo, M. M., Simola, K. O., Valanne, L. K., Larsen, A. T., Kähkönen, M., Auranen, M. P., & Ignatius, J. (2006). Pitt–Hopkins syndrome in two patients and further definition of the phenotype. *Clinical dysmorphology*, 15(2), 47-54.
- Peñagarikano, O., Abrahams, B. S., Herman, E. I., Winden, K. D., Gdalyahu, A., Dong, H., ... & Golshani, P. (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell*, 147(1), 235-246.
- Pennacchio, L. A., & Rubin, E. M. (2001). Genomic strategies to identify mammalian regulatory sequences. *Nature reviews genetics*, 2(2), 100-109.
- Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., ... & Plajzer-Frick, I. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, 444(7118), 499-502.
- Peralta, V. & Cuesta, M. J. (1995). Negative symptoms in schizophrenia: a confirmatory factor analysis of competing models. *American Journal of Psychiatry*, 152, 1450-1457.
- Peralta, V., & Cuesta, M. J. (2003). The nosology of psychotic disorders: a comparison among competing classification systems. *Schizophrenia Bulletin*, 29(3), 413.
- Pérez-Santiago, J., Diez-Alarcia, R., Callado, L. F., Zhang, J. X., Chana, G., White, C. H., ... & Woelk, C. H. (2012). A combined analysis of microarray gene expression studies of the human

prefrontal cortex identifies genes implicated in schizophrenia. *Journal of psychiatric research*, 46(11), 1464-1474.

Pergola, G., Selvaggi, P., Trizio, S., Bertolino, A., & Blasi, G. (2015). The role of the thalamus in schizophrenia from a neuroimaging perspective. *Neuroscience & Biobehavioral Reviews*, 54, 57-75.

Perkins, D. O., Jeffries, C. D., Jarskog, L. F., Thomson, J. M., Woods, K., Newman, M. A., ... & Hammond, S. M. (2007). microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. *Genome biology*, 8(2), R27.

Petrovski, S., Gussow, A. B., Wang, Q., Halvorsen, M., Han, Y., Weir, W. H., ... & Goldstein, D. B. (2015). The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. *PLoS Genet*, 11(9), e1005492.

Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., & Goldstein, D. B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*, 9(8), e1003709.

Pfriege, F. W., & Barres, B. A. (1997). Synaptic efficacy enhanced by glial cells in vitro. *Science*, 277(5332), 1684-1687.

Picardi, A., Viroli, C., Tarsitani, L., Miglio, R., de Girolamo, G., Dell'Acqua, G., & Biondi, M. (2012). Heterogeneity and symptom structure of schizophrenia. *Psychiatry research*, 198(3), 386-394.

Picchioni, M. M., Murray, R. M. (2007) Schizophrenia. *Bmj*, 335(7610): 91-95.

Pickard, B. S., Christoforou, A., Thomson, P. A., Fawkes, A., Evans, K. L., Morris, S. W., ... & Muir, W. J. (2009). Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. *Molecular psychiatry*, 14(9), 874-884.

Pickard, B. S., Malloy, M. P., Porteous, D. J., Blackwood, D. H. R., & Muir, W. J. (2005). Disruption of a brain transcription factor, NPAS3, is associated with schizophrenia and learning disability. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 136(1), 26-32.

Pieper, A. A., Wu, X., Han, T. W., Estill, S. J., Dang, Q., Wu, L. C., ... & McKnight, S. L. (2005). The neuronal PAS domain protein 3 transcription factor controls FGF-mediated adult hippocampal neurogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 14052-14057.

- Pirooznia, M., Wang, T., Avramopoulos, D., Valle, D., Thomas, G., Haganir, R. L., ...&Zandi, P. P. (2012). SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics*, 28(6), 897-899.
- Piskol, R., Ramaswami, G., & Li, J. B. (2013). Reliable identification of genomic variants from RNA-seq data. *The American Journal of Human Genetics*, 93(4), 641-651.
- Piton, A., Gauthier, J., Hamdan, F. F., Lafreniere, R. G., Yang, Y., Henrion, E., ...& Spiegelman, D. (2011). Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Molecular psychiatry*, 16(8), 867-880.
- Pitt, D., & Hopkins, I. (1978). A syndrome of mental retardation, wide mouth and intermittent overbreathing. *Australian paediatric journal*, 14(3), 182-184.
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research*, 20(1), 110-121.
- Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., ... & Rosenbloom, K. R. (2006a). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet*, 2(10), e168.
- Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., ... & Kern, A. D. (2006b). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, 443(7108), 167-172.
- Ponting, C. P., & Hardison, R. C. (2011). What fraction of the human genome is functional?. *Genome research*, 21(11), 1769-1776.
- Porteous, D. (2008). Genetic causality in schizophrenia and bipolar disorder: out with the old and in with the new. *Current opinion in genetics & development*, 18(3), 229-234.
- Potvin, S., Stip, E., Sepehry, A. A., Gendron, A., Bah, R., & Kouassi, E. (2008). Inflammatory cytokine alterations in schizophrenia: a systematic quantitative review. *Biological psychiatry*, 63(8), 801-808.
- Pouilhe, M., Gilardi-Hebenstreit, P., Dinh, C. D. T., & Charnay, P. (2007). Direct regulation of vHnf1 by retinoic acid signaling and MAF-related factors in the neural tube. *Developmental biology*, 309(2), 344-357.
- Prabakaran, S., Swatton, J. E., Ryan, M. M., Huffaker, S. J., Huang, J. J., Griffin, J. L., ... & Karp, N. A. (2004). Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Molecular psychiatry*, 9(7), 684-697.

- Prabhakar, S., Noonan, J. P., Pääbo, S., & Rubin, E. M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science*, 314(5800), 786-786.
- Prescott, C. A., & Gottesman, I. I. (1993). Genetically mediated vulnerability to schizophrenia. *Psychiatric Clinics of North America*.
- Preuss, T. M. (2012). Human brain evolution: from gene discovery to phenotype discovery. *Proceedings of the National Academy of Sciences*, 109(Supplement 1), 10709-10716.
- Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases?. *The American Journal of Human Genetics*, 69(1), 124-137.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- Pulver, A. E., Karayiorgou, M., Wolyniec, P. S., Lasseter, V. K., Kasch, L., Nestadt, G., ... & Ott, J. (1994). Sequential strategy to identify a susceptibility gene for schizophrenia: Report of potential linkage on chromosome 22q12-q13. 1: Part 1. *American journal of medical genetics*, 54(1), 36-43.
- Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., ... & Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, 506(7487), 185-190.
- Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., ... & O'Dushlaine, C. T. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256), 748-752.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559-575.
- Qi, H., Moran, M. M., Navarro, B., Chong, J. A., Krapivinsky, G., Krapivinsky, L., ... & Clapham, D. E. (2007). All four CatSper ion channel proteins are required for male fertility and sperm cell hyperactivated motility. *Proceedings of the National Academy of Sciences*, 104(4), 1219-1223.
- Quednow, B. B., Brzózka, M. M., & Rossner, M. J. (2014). Transcription factor 4 (TCF4) and schizophrenia: integrating the animal and the human perspective. *Cellular and Molecular Life Sciences*, 71(15), 2815-2835.

Bibliografía

- Quednow, B. B., Ettinger, U., Mössner, R., Rujescu, D., Giegling, I., Collier, D. A., ... & Wagner, M. (2011). The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia spectrum and healthy volunteers. *The Journal of neuroscience*, 31(18), 6684-6691.
- Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics*, 11-12.
- Rajkowska, G., & Stockmeier, C. A. (2013). Astrocyte pathology in major depressive disorder: insights from human postmortem brain tissue. *Current drug targets*, 14(11), 1225.
- Ramaswami, G., Zhang, R., Piskol, R., Keegan, L. P., Deng, P., O'Connell, M. A., & Li, J. B. (2013). Identifying RNA editing sites using RNA sequencing data alone. *Nature methods*, 10(2), 128-132.
- Ramskold, D., Wang, E. T., Burge, C. B., & Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol*, 5(12), e1000598.
- Rapoport, J. L., Giedd, J. N., & Gogtay, N. (2012). Neurodevelopmental model of schizophrenia: update 2012. *Molecular psychiatry*, 17(12), 1228-1238..
- Raz, S., Goldstein, R., Hopkins, T. L., Lauterbach, M. D., Shah, F., Porter, C. L., ... & Sander, C. J. (1994). Sex differences in early vulnerability to cerebral injury and their neurodevelopmental implications. *Psychobiology*, 22(3), 244-253.
- Reed, G. H., & Wittwer, C. T. (2004). Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clinical chemistry*, 50(10), 1748-1754.
- Reed, G. H., Kent, J. O., & Wittwer, C. T. (2007). High-resolution DNA melting analysis for simple and efficient molecular diagnostics.
- Rees, E., Walters, J. T., Georgieva, L., Isles, A. R., Chambert, K. D., Richards, A. L., ... & O'Donovan, M. C. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. *The British Journal of Psychiatry*, 204(2), 108-114.
- Rees, E., Walters, J. T., Georgieva, L., Isles, A. R., Chambert, K. D., Richards, A. L., ... & O'Donovan, M. C. (2013). Analysis of copy number variations at 15 schizophrenia-associated loci. *The British Journal of Psychiatry*, bjp-bp.

- Rehm, H. L. (2013). Disease-targeted sequencing: a cornerstone in the clinic. *Nature Reviews Genetics*, 14(4), 295-300.
- Reick, M., Garcia, J. A., Dudley, C., & McKnight, S. L. (2001). NPAS2: an analog of clock operative in the mammalian forebrain. *Science*, 293(5529), 506-509.
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, 447(7143), 425-432.
- Ren, Y., Cui, Y., Li, X., Wang, B., Na, L., Shi, J., ... & Xu, Y. (2015). A co-expression network analysis reveals lncRNA abnormalities in peripheral blood in early-onset schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 63, 1-5.
- Ren, Y., Wang, H., & Xiao, L. (2013). Improving myelin/oligodendrocyte-related dysfunction: a new mechanism of antipsychotics in the treatment of schizophrenia?. *The International Journal of Neuropsychopharmacology*, 16(03), 691-700.
- Repik, A., Rebbapragada, A., Johnson, M. S., Haznedar, J. Ö., Zhulin, I. B., & Taylor, B. L. (2000). PAS domain residues involved in signal transduction by the Aer redox sensor of *Escherichia coli*. *Molecular microbiology*, 36(4), 806-816.
- Richards, A. L., Jones, L., Moskvina, V., Kirov, G., Gejman, P. V., Levinson, D. F., ... & Owen, M. J. (2012). Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Molecular psychiatry*, 17(2), 193-201.
- Rietschel, M., Mattheisen, M., Degenhardt, F., Kahn, R. S., Linszen, D. H., van Os, J., ... & Krabbendam, L. (2012). Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Molecular psychiatry*, 17(9), 906-917.
- Riley, B., & Kendler, K. S. (2006). Molecular genetic studies of schizophrenia. *European Journal of Human Genetics*, 14(6), 669-680.
- Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., ... & Kim, Y. (2013a). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature genetics*, 45(10), 1150-1159.
- Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., ... & Heath, A. C. (2013b). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry*, 18(4), 497-511.

- Ririe, K. M., Rasmussen, R. P., & Wittwer, C. T. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical biochemistry*, 245(2), 154-160.
- Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature*, 405(6788), 847-856.
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281), 1516-1517.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.
- Roussos, P., Katsel, P., Davis, K. L., Siever, L. J., & Haroutunian, V. (2012). A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. *Archives of general psychiatry*, 69(12), 1205-1213.
- Roussos, P., Mitchell, A. C., Voloudakis, G., Fullard, J. F., Pothula, V. M., Tsang, J., ...& Okada, Y. (2014). A role for noncoding variation in schizophrenia. *Cell reports*, 9(4), 1417-1429.
- Ryan, A. K., Goodship, J. A., Wilson, D. I., Philip, N., Levy, A., Seidel, H., ... & Aurias, A. (1997). Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study. *Journal of Medical Genetics*, 34(10), 798-804.
- Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., ...& Hunt, S. E. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822), 928-933.
- Saetre, P., Emilsson, L., Axelsson, E., Kreuger, J., Lindholm, E., & Jazin, E. (2007). Inflammation-related genes up-regulated in schizophrenia brains. *Bmc Psychiatry*, 7(1), 46.
- Saha, S., Chant, D. C., Welham, J. L., & McGrath, J. J. (2006). The incidence and prevalence of schizophrenia varies with latitude. *Acta Psychiatrica Scandinavica*, 114(1), 36-39.
- Saha, S., Chant, D., & McGrath, J. (2007). A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time?. *Archives of general psychiatry*, 64(10), 1123-1131.
- Saha, S., Chant, D., Welham, J., & McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. *PLoS medicine*, 2(5), 413.

- Saiz Ruiz, J., Vega Sánchez, D. C., & Sánchez Páez, P. (2010). Bases neurobiológicas de la Esquizofrenia. *Clínica y Salud*, 21(3), 235-254.
- Salgado-Pineda, P., Junqué, C., Vendrell, P., Baeza, I., Bargalló, N., Falcón, C., & Bernardo, M. (2004). Decreased cerebral activation during CPT performance: structural and functional deficits in schizophrenic patients. *Neuroimage*, 21(3), 840-847.
- Sanacora, G., & Banasr, M. (2013). From pathophysiology to novel antidepressant drugs: glial contributions to the pathology and treatment of mood disorders. *Biological psychiatry*, 73(12), 1172-1179.
- Sandelin, A., Bailey, P., Bruce, S., Engström, P. G., Klos, J. M., Wasserman, W. W., ... & Lenhard, B. (2004). Arrays of ultraconserved non-coding regions span the *loci* of key developmental genes in vertebrate genomes. *BMC genomics*, 5(1), 99.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.
- Santarelli, D. M., Beveridge, N. J., Tooney, P. A., & Cairns, M. J. (2011). Upregulation of dicer and microRNA expression in the dorsolateral prefrontal cortex Brodmann area 46 in schizophrenia. *Biological psychiatry*, 69(2), 180-187.
- Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, 489(7414), 109-113.
- Sardi, F., Fassina, L., Venturini, L., Inguscio, M., Guerriero, F., Rolfo, E., & Ricevuti, G. (2011). Alzheimer's disease, autoimmunity and inflammation. The good, the bad and the ugly. *Autoimmunity reviews*, 11(2), 149-153.
- Satoh, J. I., Yamamoto, Y., Asahina, N., Kitano, S., & Kino, Y. (2014). RNA-Seq data mining: downregulation of NeuroD6 serves as a possible biomarker for Alzheimer's disease brains. *Disease markers*, 2014.
- Saul, M. C., Gessay, G. M., & Gammie, S. C. (2012). A new mouse model for mania shares genetic correlates with human bipolar disorder. *PLoS One*, 7(6), e38128.
- Saunders, A. M., Strittmatter, W. J., Schmechel, D., George-Hyslop, P. S., Pericak-Vance, M. A., Joo, S. H., ... & Hulette, C. (1993). Association of apolipoprotein E allele $\epsilon 4$ with late-onset familial and sporadic Alzheimer's disease. *Neurology*, 43(8), 1467-1467.

Bibliografía

- Schaaf, C. P., Gonzalez-Garay, M. L., Xia, F., Potocki, L., Gripp, K. W., Zhang, B., ... & Caskey, C. T. (2013). Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. *Nature genetics*, 45(11), 1405-1408.
- Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human molecular genetics*, 19(R2), R227-R240.
- Schanze, D., Ekici, A. B., Pfuhlmann, B., Reis, A., & Stöber, G. (2012). Evaluation of conserved and ultra-conserved non-genic sequences in chromosome 15q15-linked periodic catatonia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 159(1), 77-86.
- Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S., & Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. *Genome research*, 22(9), 1748-1759.
- Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature genetics*, 43(10), 969-976.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), 421-427.
- Schmitt, A., Steyskal, C., Bernstein, H. G., Schneider-Axmann, T., Parlapani, E., Schaeffer, E. L., ... & Falkai, P. (2009). Stereologic investigation of the posterior part of the hippocampus in schizophrenia. *Acta neuropathologica*, 117(4), 395-407.
- Schmitt, A., Wilczek, K., Blennow, K., Maras, A., Jatzko, A., Petroianu, G., ... & Gattaz, W. F. (2004). Altered thalamic membrane phospholipids in schizophrenia: a postmortem study. *Biological psychiatry*, 56(1), 41-45.
- Schneider, K. (1959). Clinical psychopathology. (Trans. by MW Hamilton).
- Schork, A. J., Thompson, W. K., Pham, P., Torkamani, A., Roddey, J. C., Sullivan, P. F., ... & Andreassen, O. A. (2013). All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet*, 9(4), e1003449.
- Schuettengruber, B., & Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development*, 136(21), 3531-3542.

- Schwab, C., Yu, S., Wong, W., McGeer, E. G., & McGeer, P. L. (2013). GAD65, GAD67, and GABAT immunostaining in human brain and apparent GAD65 loss in Alzheimer's disease. *Journal of Alzheimer's Disease*, 33(4), 1073-1088.
- Schwartz, Y. B., & Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nature Reviews Genetics*, 14(12), 853-864.
- Schwartz, Y. B., Kahn, T. G., Nix, D. A., Li, X. Y., Bourgon, R., Biggin, M., & Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature genetics*, 38(6), 700-705.
- Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., ... & Leotta, A. (2007). Strong association of de novo copy number mutations with autism. *Science*, 316(5823), 445-449.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., ...& Navin, N. (2004). Large-scale copy number polymorphism in the human genome. *Science*, 305(5683), 525-528.
- Sebat, J., Levy, D. L., & McCarthy, S. E. (2009). Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends in Genetics*, 25(12), 528-535.
- Seeman, M. V. (1996). The role of estrogen in schizophrenia. *Journal of Psychiatry and Neuroscience*, 21(2), 123.
- Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., ... & Genovese, G. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*, 530(7589), 177-183.
- Sekar, S., McDonald, J., Cuyugan, L., Aldrich, J., Kurdoglu, A., Adkins, J., ...&Liang, W. S. (2015). Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes. *Neurobiology of aging*, 36(2), 583-591.
- Sepp, M., Kannike, K., Eesmaa, A., Urb, M., & Timmusk, T. (2011). Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. *PloS one*, 6(7), e22138.
- Sequeira, P. A., Martin, M. V., & Vawter, M. P. (2012). The first decade and beyond of transcriptional profiling in schizophrenia. *Neurobiology of disease*, 45(1), 23-36.
- Sequencing, T. C., & Analysis Consortium. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055), 69-87.

Bibliografía

Shakiba, A. (2014). The role of the cerebellum in neurobiology of psychiatric disorders. *Neurologic clinics*, 32(4), 1105-1115.

Sham, P., Bader, J. S., Craig, I., O'Donovan, M., & Owen, M. (2002). DNA pooling: a tool for large-scale association studies. *Nature Reviews Genetics*, 3(11), 862-871.

Sharma, R. P., Grayson, D. R., & Gavin, D. P. (2008). Histone deacetylase 1 expression is increased in the prefrontal cortex of schizophrenia subjects: analysis of the National Brain Databank microarray collection. *Schizophrenia research*, 98(1), 111-117.

Sharma, R. P., Janicak, P. G., Bissette, G., & Nemeroff, C. B. (1997). CSF neurotensin concentrations and antipsychotic treatment in schizophrenia and schizoaffective disorder. *American Journal of Psychiatry*, 154(7), 1019-1021.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, 29(1), 308-311.

Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., ... & Olincy, A. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, 460(7256), 753-757.

Shifman, S., Johannesson, M., Bronstein, M., Chen, S. X., Collier, D. A., Craddock, N. J., ... & Owen, M. J. (2008). Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women.

Shifman, S., Kuypers, J., Kokoris, M., Yakir, B., & Darvasi, A. (2003). Linkage disequilibrium patterns of the human genome across populations. *Human molecular genetics*, 12(7), 771-776.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., ... & Fukuda, S. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences*, 100(26), 15776-15781.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., ... & Weinstock, G. M. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research*, 15(8), 1034-1050.

Simon, J. A., & Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature reviews Molecular cell biology*, 10(10), 697-708.

Sklar, P., Ripke, S., Scott, L. J., Andreassen, O. A., Cichon, S., Craddock, N., ... & Corvin, A. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics*, 43(10), 977.

Smith, N. G., Brandström, M., & Ellegren, H. (2004). Evidence for turnover of functional noncoding DNA in mammalian genome evolution. *Genomics*, 84(5), 806-813.

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., ...& Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature*, 511(7511), 611-615.

Sofroniew, M. V., & Vinters, H. V. (2010). Astrocytes: biology and pathology. *Acta neuropathologica*, 119(1), 7-35.

Sokal, R. R., & Rohlf, F. J. (1969). *The principles and practice of statistics in biological research* (pp. 399-400). San Francisco:: WH Freeman and company.

Sokolowski, M., Wasserman, J., & Wasserman, D. (2015). Polygenic associations of neurodevelopmental genes in suicide attempt. *Molecular psychiatry*.

Song, C., Merali, Z., & Anisman, H. (1999). Variations of nucleus accumbens dopamine and serotonin following systemic interleukin-1, interleukin-2 or interleukin-6 treatment. *Neuroscience*, 88(3), 823-836.

Song, W., Li, W., Feng, J., Heston, L. L., Scaringe, W. A., & Sommer, S. S. (2008). Identification of high risk DISC1 structural variants with a 2% attributable risk for schizophrenia. *Biochemical and biophysical research communications*, 367(3), 700-706.

Sørensen, H. J., Mortensen, E. L., Reinisch, J. M., & Mednick, S. A. (2004). Association between prenatal exposure to analgesics and risk of schizophrenia. *The British Journal of Psychiatry*, 185(5), 366-371.

Spadaro, P. A., Flavell, C. R., Widagdo, J., Ratnu, V. S., Troup, M., Ragan, C., ... & Bredy, T. W. (2015). Long noncoding RNA-directed epigenetic regulation of gene expression is associated with anxiety-like behavior in mice. *Biological psychiatry*, 78(12), 848-859.

Squeglia, L. M., Rinker, D. A., Bartsch, H., Castro, N., Chung, Y., Dale, A. M., ...& Tapert, S. F. (2014). Brain volume reductions in adolescent heavy drinkers. *Developmental cognitive neuroscience*, 9, 117-125.

St Clair, D., Xu, M., Wang, P., Yu, Y., Fang, Y., Zhang, F., ...&He, L. (2005). Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961. *Jama*, 294(5), 557-562.

Stearns, S. C. (1992). *The evolution of life histories* (Vol. 249). Oxford: Oxford University Press.

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., ...& Sigurdsson, E. (2009). Common variants conferring risk of schizophrenia. *Nature*, 460(7256), 744-747.

Stefansson, H., Petursson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., ...& Chou, T. T. (2002). Neuregulin 1 and susceptibility to schizophrenia. *The American Journal of Human Genetics*, 71(4), 877-892.

Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O. P., Ingason, A., Steinberg, S., ... & Hansen, T. (2008). Large recurrent microdeletions associated with schizophrenia. *nature*, 455(7210), 232-236.

Steinberg, S., de Jong, S., Andreassen, O. A., Werge, T., Børglum, A. D., Mors, O., ...& Demontis, D. (2011a). Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Human molecular genetics*, 20(20), 4076-4081.

Steinberg, S., de Jong, S., Mattheisen, M., Costas, J., Demontis, D., Jamain, S., ... & Sigurdsson, E. (2014). Common variant at 16p11. 2 conferring risk of psychosis. *Molecular psychiatry*, 19(1), 108-114.

Steinberg, S., Mors, O., Børglum, A. D., Gustafsson, O., Werge, T., Mortensen, P. B., ... & Olason, P. (2011b). Expanding the range of ZNF804A variants conferring risk of psychosis. *Molecular psychiatry*, 16(1), 59-66.

Stephens, S. H., Logel, J., Barton, A., Franks, A., Schultz, J., Short, M., ... & Hodgkinson, C. (2009). Association of the 5'-upstream regulatory region of the $\alpha 7$ nicotinic acetylcholine receptor subunit gene (CHRNA7) with schizophrenia. *Schizophrenia research*, 109(1), 102-112.

Stevens, C. D., Altshuler, L. L., Bogerts, B., & Falkai, P. (1988). Quantitative study of gliosis in schizophrenia and Huntington's chorea. *Biological psychiatry*, 24(6), 697-700.

Straub, R. E., MacLean, C. J., O'Neill, F. A., Burke, J., Murphy, B., Duke, F., ...& Kendler, K. S. (1995). A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. *Nature genetics*, 11(3), 287-293.

- Strous, R. D., Bark, N., Woerner, M., & Lachman, H. M. (1997). Lack of association of functional catechol-O-methyl transferase gene polymorphism in schizophrenia. *Biological psychiatry*, 41(4), 493-495.
- Sturgeon, K., Kaneko, T., Biemann, M., Gauthier, A., Chawengsaksophak, K., & Cordes, S. P. (2011). Cdx1 refines positional identity of the vertebrate hindbrain by directly repressing *Mafk* expression. *Development*, 138(1), 65-74.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), 15545-15550.
- Sullivan, P. F. (2005). The genetics of schizophrenia. *PLoS medicine*, 2(7), 614.
- Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of general psychiatry*, 60(12), 1187-1192.
- Sullivan, P. F., Lin, D., Tzeng, J. Y., van den Oord, E. J. C. G., Perkins, D., Stroup, T. S., ... & Liu, W. (2008). Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Molecular psychiatry*, 13(6), 570-584.
- Sun, Z., & Hopkins, N. (2001). *vhnf1*, the *MODY5* and familial GCKD-associated gene, regulates regional specification of the zebrafish gut, pronephros, and hindbrain. *Genes & development*, 15(23), 3217-3229.
- Susser, E. S., & Lin, S. P. (1992). Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944-1945. *Archives of general psychiatry*, 49(12), 983-988.
- Sutherland, R. J., Whishaw, I. Q., & Kolb, B. (1988). Contributions of cingulate cortex to two forms of spatial learning and memory. *The Journal of neuroscience*, 8(6), 1863-1872.
- Svrakic, D. M., Zorumski, C. F., Svrakic, N. M., Zwir, I., & Cloninger, C. R. (2013). Risk architecture of schizophrenia: the role of epigenetics. *Current opinion in psychiatry*, 26(2), 188-195.
- Syvänen, A. C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2(12), 930-942.

Bibliografía

- Szatkiewicz, J. P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J. L., Neale, B. M., ...& Kähler, A. (2014). Copy number variation in schizophrenia in Sweden. *Molecular psychiatry*, 19(7), 762-773.
- Szulwach, K. E., Li, X., Smrt, R. D., Li, Y., Luo, Y., Lin, L., ... & Jin, P. (2010). Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *The Journal of cell biology*, 189(1), 127-141.
- Tabarki, B., Al Mutairi, F., & Al Hashem, A. (2014). The fragile site WWOX gene and the developing brain. *Experimental Biology and Medicine*, 1535370214561952.
- Takahashi, N., & Sakurai, T. (2013). Roles of glial cells in schizophrenia: possible targets for therapeutic approaches. *Neurobiology of disease*, 53, 49-60.
- Takahashi, N., Sakurai, T., Davis, K. L., & Buxbaum, J. D. (2011). Linking oligodendrocyte and myelin dysfunction to neurocircuitry abnormalities in schizophrenia. *Progress in neurobiology*, 93(1), 13-24.
- Takahashi, T., Suzuki, M., Kawasaki, Y., Hagino, H., Yamashita, I., Nohara, S., ... & Kurachi, M. (2003). Perigenual cingulate gyrus volume in patients with schizophrenia: a magnetic resonance imaging study. *Biological psychiatry*, 53(7), 593-600.
- Takata, A., Xu, B., Ionita-Laza, I., Roos, J. L., Gogos, J. A., & Karayiorgou, M. (2014). Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron*, 82(4), 773-780.
- Tan, M. L., Dyck, B. A., Gabriele, J., Daya, R. P., Thomas, N., Sookram, C., ... & Mishra, R. K. (2014). Synapsin II gene expression in the dorsolateral prefrontal cortex of brain specimens from patients with schizophrenia and bipolar disorder: effect of lifetime intake of antipsychotic drugs. *The pharmacogenomics journal*, 14(1), 63-69.
- Tandon, R., & Greden, J. F. (1989). Cholinergic hyperactivity and negative schizophrenic symptoms: a model of cholinergic/dopaminergic interactions in schizophrenia. *Archives of General Psychiatry*, 46(8), 745-753.
- Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R. E., Heckers, S., ...& Van Os, J. (2013). Definition and description of schizophrenia in the DSM-5. *Schizophrenia research*, 150(1), 3-10.
- Taylor, B. L., & Zhulin, I. B. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews*, 63(2), 479-506.

- The International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455(7210), 237-241.
- Thevenon, J., Lopez, E., Keren, B., Heron, D., Mignot, C., Altuzarra, C., ... & Minot, D. (2012). Intragenic CAMTA1 rearrangements cause non-progressive congenital ataxia with or without intellectual disability. *Journal of medical genetics*, 49(6), 400-408.
- Thomas, M. G., Welch, C., Stone, L., Allan, P., Barker, R. A., & White, R. B. (2015). PAX6 expression may be protective against dopaminergic cell loss in Parkinson's disease. *CNS & neurological disorders drug targets*.
- Thompson, J. L., Pogue-Geile, M. F., & Grace, A. A. (2004). Developmental pathology, dopamine, and stress: a model for the age of onset of schizophrenia symptoms. *Schizophrenia bulletin*, 30(4), 875-900.
- Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., ... & Turner, D. J. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, 464(7291), 1082-1086.
- Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., ... & Garg, K. (2012). The accessible chromatin landscape of the human genome. *Nature*, 489(7414), 75-82.
- Tienari, P., Wynne, L. C., Moring, J., & Lahti, I. (1994). The Finnish adoptive family study of schizophrenia: Implications for family research. *The British Journal of Psychiatry*.
- Timms, A. E., Dorschner, M. O., Wechsler, J., Choi, K. Y., Kirkwood, R., Girirajan, S., ... & Horwitz, M. S. (2013). Support for the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. *JAMA psychiatry*, 70(6), 582-590.
- Tishkoff, S. A., & Verrelli, B. C. (2003). Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annual review of genomics and human genetics*, 4(1), 293-340.
- Tkachev, D., Mimmack, M. L., Ryan, M. M., Wayland, M., Freeman, T., Jones, P. B., ... & Bahn, S. (2003). Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *The Lancet*, 362(9386), 798-805.
- Tolosa, A., Sanjuán, J., Leal, C., Costas, J., Moltó, M. D., & De Frutos, R. (2008). Rapid evolving RNA gene HAR1A and schizophrenia. *Schizophrenia research*, 99(1), 370-372.

- Torkamani, A., Dean, B., Schork, N. J., & Thomas, E. A. (2010). Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Research*, 20(4), 403-412.
- Torres, G., & Horowitz, J. M. (1999). Drugs of abuse and brain gene expression. *Psychosomatic medicine*, 61(5), 630-650.
- Torrey, E. F., Barci, B. M., Webster, M. J., Bartko, J. J., Meador-Woodruff, J. H., & Knable, M. B. (2005). Neurochemical markers for schizophrenia, bipolar disorder, and major depression in postmortem brains. *Biological psychiatry*, 57(3), 252-260.
- Tovilla-Zarate, C., Medellin, B. C., Fresan, A., Apiquian, R., Dassori, A., Rolando, M., ... & Nicolini, H. (2009). APOE-ε3 and APOE-219G Haplotypes Increase the Risk for Schizophrenia in Sibling Pairs. *The Journal of neuropsychiatry and clinical neurosciences*.
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), 1105-1111.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3), 562-578.
- Trerotola, M., Relli, V., Simeone, P., & Alberti, S. (2015). Epigenetic inheritance and the missing heritability. *Human genomics*, 9(1), 1.
- Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., ... & Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 329(5992), 689-693.
- Uittenbogaard, M., Baxter, K. K., & Chiaramello, A. (2010). The neurogenic basic helix-loop-helix transcription factor NeuroD6 confers tolerance to oxidative stress by triggering an antioxidant response and sustaining the mitochondrial biomass. *ASN neuro*, 2(2), AN20100005.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research*, 40(15), e115-e115.
- Vachev, T. I., Popov, N. T., Stoyanova, V. K., Ivanov, H. Y., & Minchev, D. S. (2016). Down Regulation of MIR-320 Gene Family Members in the Peripheral Blood of Schizophrenia Patients. *Int. J. Curr. Microbiol. App. Sci*, 5(1), 221-230.

- Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H. H., Peoples, A., ...& Iakoucheva, L. M. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature*, 471(7339), 499-503.
- van der Werf, M., Hanssen, M., Köhler, S., Verkaaik, M., Verhey, F. R., van Winkel, R., ... & Allardyce, J. (2014). Systematic review and collaborative recalculation of 133 693 incident cases of schizophrenia. *Psychological medicine*, 44(01), 9-16.
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in genetics*, 30(9), 418-426.
- van Os, J., Kenis, G., & Rutten, B. P. (2010). The environment and schizophrenia. *Nature*, 468(7321), 203-212.
- Varambally, S., Dhanasekaran, S. M., Zhou, M., Barrette, T. R., Kumar-Sinha, C., Sanda, M. G., ... & Rubin, M. A. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*, 419(6907), 624-629.
- Veijola, J., Guo, J. Y., Moilanen, J. S., Jääskeläinen, E., Miettunen, J., Kyllönen, M., ...&Kiviniemi, V. (2014). Longitudinal changes in total brain volume in schizophrenia: relation to symptom severity, cognition and antipsychotic medication.
- Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, 270(5235), 484-487.
- Veldic, M., Caruncho, H. J., Liu, W. S., Davis, J., Satta, R., Grayson, D. R., ... & Costa, E. (2004). DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains. *Proceedings of the National Academy of Sciences*, 101(1), 348-353.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Gocayne, J. D. (2001). The sequence of the human genome. *science*, 291(5507), 1304-1351.
- Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., ...& Eussen, B. E. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, 65(5), 905-914.
- Vernon, A. C., Crum, W. R., Lerch, J. P., Chege, W., Natesan, S., Modo, M., ... & Kapur, S. (2014). Reduced cortical volume and elevated astrocyte density in rats chronically treated with antipsychotic drugs—linking magnetic resonance imaging findings to cellular pathology. *Biological psychiatry*, 75(12), 982-990.

- Vilella, E., Costas, J., Sanjuan, J., Guitart, M., De Diego, Y., Carracedo, A., ...&Najera, C. (2008). Association of schizophrenia with DTNBP1 but not with DAO, DAOA, NRG1 and RGS4 nor their genetic interaction. *Journal of psychiatric research*, 42(4), 278-288.
- Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007). VISTA Enhancer Browser-a database of tissue-specific human enhancers. *Nucleic Acids Res* 35:D88-92
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., ... & Afzal, V. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), 854-858.
- Visscher, P. M., Medland, S. E., Ferreira, M. A., Morley, K. I., Zhu, G., Cornes, B. K., ...& Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*, 2(3), e41.
- Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*, 55(4), 641-658.
- Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., ...&Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 474(7351), 380-384.
- Volterra, A., & Meldolesi, J. (2005). Astrocytes, from brain glue to communication elements: the revolution continues. *Nature Reviews Neuroscience*, 6(8), 626-640.
- Vossen, R. H., Aten, E., Roos, A., & den Dunnen, J. T. (2009). High-Resolution Melting Analysis (HRMA)—More than just sequence variant screening. *Human mutation*, 30(6), 860-866.
- Vuong, H., Stephens, R. M., & Volfovsky, N. (2012). AVIA: an interactive web-server for annotation, visualization and impact analysis of genomic variations. In *BMC Proceedings* (Vol. 6, No. Suppl 6, p. P37). BioMed Central Ltd.
- Wacholder, S., Chanock, S., Garcia-Closas, M., & Rothman, N. (2004).Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, 96(6), 434-442.
- Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., ...& Stray, S. M. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *science*, 320(5875), 539-543.
- Wang, D., Zhou, Y., Zhuo, C., Qin, W., Zhu, J., Liu, H., ... & Yu, C. (2015). Altered functional connectivity of the cingulate subregions in schizophrenia. *Translational psychiatry*, 5(6), e575.

- Wang, J. F., Shao, L., Sun, X., & Young, L. T. (2009a). Increased oxidative stress in the anterior cingulate cortex of subjects with bipolar disorder and schizophrenia. *Bipolar disorders*, 11(5), 523-529.
- Wang, J., Zhuang, J., Iyer, S., Lin, X. Y., Greven, M. C., Kim, B. H., ... & Birney, E. (2013). Factorbook. org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic acids research*, 41(D1), D171-D176.
- Wang, K., Li, M., & Hakonarson, H. (2010a). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16), e164-e164.
- Wang, K., Li, M., & Hakonarson, H. (2010b). Analysing biological pathways in genome-wide association studies. *Nature Reviews Genetics*, 11(12), 843-854.
- Wang, S., Sun, C. E., Walczak, C. A., Ziegler, J. S., Kipps, B. R., Goldin, L. R., & Diehl, S. R. (1995). Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. *Nature genetics*, 10(1), 41-46.
- Wang, X. (2014). Firth logistic regression for rare variant association tests. *Frontiers in genetics*, 5, 187.
- Wang, Z., Gerstein, M., & Snyder, M. (2009b). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63.
- Ward, L. D., & Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research*, 40(D1), D930-D934.
- Warnica, W., Merico, D., Costain, G., Alfred, S. E., Wei, J., Marshall, C. R., ... & Bassett, A. S. (2015). Copy number variable microRNAs in schizophrenia and their neurodevelopmental gene targets. *Biological psychiatry*, 77(2), 158-166.
- Weber, H., Kittel-Schneider, S., Gessner, A., Domschke, K., Neuner, M., Jacob, C. P., ... & Baune, B. T. (2011). Cross-disorder analysis of bipolar risk genes: further evidence of DGKH as a risk gene for bipolar disorder, but also unipolar depression and adult ADHD. *Neuropsychopharmacology*, 36(10), 2076-2085.
- Weidenhofer, J., Bowden, N. A., Scott, R. J., & Tooney, P. A. (2006). Altered gene expression in the amygdala in schizophrenia: up-regulation of genes located in the cytomatrix active zone. *Molecular and Cellular Neuroscience*, 31(2), 243-250.

Bibliografía

- Weinberger, D. R. (1986). The pathogenesis of schizophrenia: a neurodevelopmental theory. *Handbook of schizophrenia*, 1, 397-406.
- Weinberger, D. R. (1987). Implications of normal brain development for the pathogenesis of schizophrenia. *Archives of general psychiatry*, 44(7), 660-669.
- Weinhold, N., Jacobsen, A., Schultz, N., Sander, C., & Lee, W. (2014). Genome-wide analysis of noncoding regulatory mutations in cancer. *Nature genetics*, 46(11), 1160-1165.
- Weiss, K. M., & Clark, A. G. (2002). Linkage disequilibrium and the mapping of complex human traits. *TRENDS in Genetics*, 18(1), 19-24.
- Weiss, K. M., & Terwilliger, J. D. (2000). How many diseases does it take to map a gene with SNP?. *Nature genetics*, 26(2), 151-158.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., ... & Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*, 42(D1), D1001-D1006.
- Wenger, A. M., Clarke, S. L., Notwell, J. H., Chung, T., Tuteja, G., Guturu, H., ... & Bejerano, G. (2013). The enhancer landscape during early neocortical development reveals patterns of dense regulation and co-option. *PLoS Genet*, 9, e1003728.
- Whalen, S., Héron, D., Gaillon, T., Moldovan, O., Rossi, M., Devillard, F., ... & Charles, P. (2012). Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: clinical score and further delineation of the TCF4 mutational spectrum. *Human mutation*, 33(1), 64-72.
- Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., ... & Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 453(7199), 1239-1243.
- Wilkinson, F. H., Park, K., & Atchison, M. L. (2006). Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proceedings of the National Academy of Sciences*, 103(51), 19296-19301.
- Willemsen, M. H., Vallès, A., Kirkels, L. A., Mastebroek, M., Loohuis, N. O., Kos, A., ... & Holder-Espinasse, M. (2011). Chromosome 1p21. 3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. *Journal of medical genetics*, jmedgenet-2011.
- Williams, H. J., Moskvina, V., Smith, R. L., Dwyer, S., Russo, G., Owen, M. J., & O'Donovan, M. C. (2011a). Association between TCF4 and schizophrenia does not exert its effect by

common nonsynonymous variation or by influencing cis-acting regulation of mRNA expression in adult human brain. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 156(7), 781-784.

Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., ... & Giegling, I. (2011b). Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Molecular psychiatry*, 16(4), 429-441.

Williams, M. R., Hampton, T., Pearce, R. K., Hirsch, S. R., Ansorge, O., Thom, M., & Maier, M. (2013). Astrocyte decrease in the subgenual cingulate and callosal genu in schizophrenia. *European archives of psychiatry and clinical neuroscience*, 263(1), 41-52.

Wirgenes, K. V., S nderby, I. E., Haukvik, U. K., Mattingsdal, M., Tesli, M., Athanasiu, L., ... & Agartz, I. (2012). TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. *Translational psychiatry*, 2(5), e112.

Witcher, M. R., Kirov, S. A., & Harris, K. M. (2007). Plasticity of perisynaptic astroglia during synaptogenesis in the mature rat hippocampus. *Glia*, 55(1), 13-23.

Wittwer, C. T. (2009). High-resolution DNA melting analysis: advancements and limitations. *Human mutation*, 30(6), 857-859.

Wong, J., Duncan, C. E., Beveridge, N. J., Webster, M. J., Cairns, M. J., & Weickert, C. S. (2012). Expression of NPAS3 in the human cortex and evidence of its posttranscriptional regulation by miR-17 during development, with implications for schizophrenia. *Schizophrenia bulletin*, sbr177.

Woodbury-Fari a, M. A. (2014). The Importance of Glia in Dealing with Stress. *Psychiatric Clinics of North America*, 37(4), 679-705.

Woolf, R., Carroll, L., Ibrahim, Y., Owen, M., & O'Donovan, M. (2013). 2788–Large-scale mutation screening of the putative autism susceptibility gene SCN2A in schizophrenia. *European Psychiatry*, 28, 1.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., ... & Walter, K. (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, 3(1), e7.

Wray, N., & Visscher, P. (2008). Estimating trait heritability. *Nature Education*, 1(1), 29.

Wu, H., Xu, J., Pang, Z. P., Ge, W., Kim, K. J., Blanchi, B., ... & Sun, Y. E. (2007). Integrative genomic and functional analyses reveal neuronal subtype differentiation bias in human

embryonic stem cell lines. *Proceedings of the National Academy of Sciences*, 104(34), 13821-13826.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, 89(1), 82-93.

Wurtman, R. J., Cansev, M., & Ulus, I. H. (2009). Choline and its products acetylcholine and phosphatidylcholine. In *Handbook of neurochemistry and molecular neurobiology* (pp. 443-501). Springer US.

Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodrick, S., Sun, Y., ...& Karayiorgou, M. (2012a). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature genetics*, 44(12), 1365-1369.

Xu, B., Roos, J. L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., ...& Karayiorgou, M. (2011a). Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nature genetics*, 43(9), 864-868.

Xu, B., Roos, J. L., Levy, S., Van Rensburg, E. J., Gogos, J. A., & Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature genetics*, 40(7), 880-885.

Xu, J., Sun, J., Chen, J., Wang, L., Li, A., Helm, M., ... & Chen, X. (2012b). RNA-Seq analysis implicates dysregulation of the immune system in schizophrenia. *BMC genomics*, 13(Suppl 8), S2.

Xu, K., Schadt, E. E., Pollard, K. S., Roussos, P., & Dudley, J. T. (2015). Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Molecular biology and evolution*, 32(5), 1148-1160.

Xu, T.Y. (2011b). The subtypes of schizophrenia. *Shanghai Arch. Psychiatr.* 23, 106–108.

Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A., & Dougherty, J. D. (2014). Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *The Journal of Neuroscience*, 34(4), 1420-1431.

Yamada, K., Iwayama, Y., Hattori, E., Iwamoto, K., Toyota, T., Ohnishi, T., ...& Yoshikawa, T. (2011). Genome-wide association study of schizophrenia in Japanese population. *PloS one*, 6(6), e20468.

- Yang, J., Visscher, P. M., & Wray, N. R. (2010). Sporadic cases are the norm for complex disease. *European Journal of Human Genetics*, 18(9), 1039-1043.
- Ye, T., Lipska, B. K., Tao, R., Hyde, T. M., Wang, L., Li, C., ... & Weinberger, D. R. (2012). Analysis of copy number variations in brain DNA from patients with schizophrenia and other psychiatric disorders. *Biological psychiatry*, 72(8), 651-654.
- Yip, K. Y., Cheng, C., Bhardwaj, N., Brown, J. B., Leng, J., Kundaje, A., ... & Gerstein, M. (2012). Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol*, 13(9), R48.
- Yu, L., Arbez, N., Nucifora, L. G., Sell, G. L., Delisi, L. E., Ross, C. A., ... & Nucifora Jr, F. C. (2014). A mutation in NPAS3 segregates with mental illness in a small family. *Molecular psychiatry*, 19(1), 7.
- Yung, A. R., & McGorry, P. D. (1996). The prodromal phase of first-episode psychosis: past and current conceptualizations. *Schizophrenia bulletin*, 22(2), 353-370.
- Zammit, S., Odd, D., Horwood, J., Thompson, A., Thomas, K., Menezes, P., ... & Harrison, G. (2009). Investigating whether adverse prenatal and perinatal events are associated with non-clinical psychotic symptoms at age 12 years in the ALSPAC birth cohort. *Psychological medicine*, 39(09), 1457-1467.
- Zeng, J., Konopka, G., Hunt, B. G., Preuss, T. M., Geschwind, D., & Soojin, V. Y. (2012). Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *The American Journal of Human Genetics*, 91(3), 455-465.
- Zhang, B., Guan, F., Chen, G., Lin, H., Zhang, T., Feng, J., ... & Fu, D. (2015). Common variants in SLC1A2 and schizophrenia: Association and cognitive function in patients with schizophrenia and healthy individuals. *Schizophrenia research*, 169(1), 128-134.
- Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keefe, S., ... & Wu, J. Q. (2014a). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *The Journal of Neuroscience*, 34(36), 11929-11947.
- Zhang, Z. H., Jhaveri, D. J., Marshall, V. M., Bauer, D. C., Edson, J., Narayanan, R. K., ... & Zhao, Q. Y. (2014b). A comparative study of techniques for differential expression analysis on RNA-Seq data.

Bibliografía

- Zhao, L., Li, J., Ma, Y., Wang, J., Pan, W., Gao, K., ... & Zhao, S. (2015). Ezh2 is involved in radial neuronal migration through regulating Reelin expression in cerebral cortex. *Scientific reports*, 5.
- Zhao, Y., Kwan, K. M., Mailloux, C. M., Lee, W. K., Grinberg, A., Wurst, W., ... & Westphal, H. (2007). LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. *Proceedings of the National Academy of Sciences*, 104(32), 13182-13186.
- Zhou, V. W., Goren, A., & Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews Genetics*, 12(1), 7-18.
- Zhou, Y. D., Barnard, M., Tian, H., Li, X., Ring, H. Z., Francke, U., ... & McKnight, S. L. (1997). Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. *Proceedings of the National Academy of Sciences*, 94(2), 713-718.
- Zondervan, K. T., & Cardon, L. R. (2004). The complex interplay among factors that influence allelic association. *Nature Reviews Genetics*, 5(2), 89-100.
- Zweier, C., Peippo, M. M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., ... & Devriendt, K. (2007). Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *The American Journal of Human Genetics*, 80(5), 994-1001.